



## 저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

체육학 석사학위논문

# 프로야구 선수의 가치 평가 요인 탐색

: 한국프로야구리그와 메이저리그의  
연봉 결정요인 비교 분석

2018년 2월

서울대학교 대학원  
체육교육과  
황 수 응

# 프로야구 선수의 가치 평가 요인 탐색

: 한국프로야구리그와 메이저리그의  
연봉 결정요인 비교 분석

지도교수 박 일 혁

이 논문을 체육학 석사학위논문으로 제출함  
2017년 12월

서울대학교 대학원  
체육교육과  
황 수 응

황수응의 석사학위논문을 인준함  
2017년 12월

위 원 장      임    충    훈      (인)

부위원장      김    유    겸      (인)

위      원      박    일    혁      (인)

# 국문초록

## 프로야구 선수의 가치 평가 요인 탐색 : 한국프로야구리그와 메이저리그의 연봉 결정요인 비교 분석

황 수 응  
서울대학교 대학원  
체육교육과

본 연구의 목적은 한국프로야구(KBO)리그와 메이저리그(MLB) 선수들의 연봉 결정요인을 탐색하고, 분석 결과를 바탕으로 KBO리그의 연봉 시장을 진단하여, 리그 특성에 맞는 연봉 책정 지표 개발을 위한 기초자료를 제시하는 것이다.

시기에 따른 한국프로야구리그의 연봉 결정요인 변화를 탐색하고자 2012~2016, 2004~2006시즌 KBO리그 출전 선수 중 규정이닝과 규정타석을 충족한 투수와 타자의 기록으로 주성분분석과 다중회귀분석을 실시하였고, 결과를 바탕으로 시기에 따른 연봉 결정요인의 변화를 비교 분석하였다. 또한 리그에 따른 연봉 결정요인을 비교 분석하기 위하여 2014~2016시즌 MLB 선수들을 대상으로 KBO리그 분석과 같은 절차의 분석을 진행하여 연봉 결정요인을 비교 분석하였다. 이 과정에서 경기력을 나타내는 다양한 세이버메트릭스 지수의 차원 축소를 위하여 주성분분석을 실시하였고, 산출된 주성분벡터는 다중회귀모형의 독립변수로 설정하였다.

본 연구를 통해 도출된 결론은 다음과 같다. 첫째, KBO리그 타자의 연봉 결정요인을 탐색한 결과 시기에 따른 변동을 보였다. 2012~2016시즌에는 타자의 경기력과 FA구분, 올스타 출전 횟수가 중요한 요인으로 나타났고, 포지션 구분, 팀 구분 또한 연봉과 유의한 관계를 보였다. 반면 2004~2006시즌 타자의 연봉은 2012~2016시즌과 달리 경기력은 유의성을 띠지 않았고, FA유무, 올스타 출전 횟수 등 특정 변수만이 중요한 요인으로 나타났다. 이와 같은 결과는 동일한 리그 내에서도 시간의 흐름에 따라 선수의 가치를 평가하는 기준과 각 요인의 중요도가 달라질 수 있음을 시사한다. 또한 2014~2016시즌 MLB 타자의 연봉 결정요인은 경기력, 경력, 몸무게, FA유무, 포지션 구분 등 투입된 대부분의 독립변수에서 유의성을 나타냈고, 특히 선수의 경력이 중요한 요인으로 나타났다. 이는 KBO리그와는 달리 세계 최고 수준인 MLB에서의 경력은, 경기력 수준이 꾸준히 유지되어야 가능한 부분이기 때문에 선수의 가치를 평가함에 있어 반영된 결과로 보인다.

둘째, KBO리그 투수의 연봉 결정요인을 탐색한 결과 시기에 따른 변동을 보였다. 2012~2016시즌에는 승리 관련 기록이 중요시되었으나, 2004~2006시즌에는 투구 능력을 나타내는 기록이 중요한 가치 평가 요인으로 나타났다. 또한 2014~2016시즌 MLB 투수의 연봉은 투구 능력에 대한 기록, 경력, 올스타 출전 횟수, FA구분, 좌투우투 구분이 연봉에 영향을 미치는 것 나타났다. 특히 KBO리그에서는 나타나지 않았던 좌완투수에 대한 긍정적인 가치 평가가 MLB에서는 나타났는데, 이와 같은 결과는 리그 특성에 따라서 선수의 가치 평가가 다르게 이루어질 수 있음을 시사하는 결과이다.

---

**주요어:** 한국프로야구, 메이저리그, 연봉, 가치 평가, 주성분분석, 다중회귀분석

**학 번 :** 2016-21636

# 목 차

I. 서 론 .....	1
1. 연구의 필요성 .....	1
2. 연구목적 .....	6
3. 용어의 정의 .....	7
II. 이론적 배경 .....	8
1. 야구에서의 경기력 지표 .....	8
1) 투수의 세이버메트릭스 지수 .....	9
2) 타자의 세이버메트릭스 지수 .....	12
2. 프로스포츠 연봉 결정요인 분석 방법 .....	15
1) 주성분분석을 활용한 연봉 결정요인 탐색 .....	15
2) 선형회귀분석을 활용한 연봉 결정요인 탐색 .....	17
3. 프로스포츠 연봉 결정요인 관련 선행 연구 .....	19

III. 연구 방법 .....	22
1. 연구 대상 .....	23
2. 연구 도구 .....	24
3. 자료의 구성.....	24
4. 자료처리방법 .....	26
IV. 연구 결과 .....	27
1. 타자의 연봉 결정요인 분석.....	29
1) 2012 ~ 2016시즌 KBO리그 타자의 연봉 결정요인 .....	29
2) 2004 ~ 2006시즌 KBO리그 타자의 연봉 결정요인 .....	35
3) 2014 ~ 2016시즌 MLB 타자의 연봉 결정요인 .....	39
2. 투수의 연봉 결정요인 분석.....	44
1) 2012 ~ 2016시즌 KBO리그 투수의 연봉 결정요인 .....	44
2) 2004 ~ 2006시즌 KBO리그 투수의 연봉 결정요인 .....	50
3) 2014 ~ 2016시즌 MLB 투수의 연봉 결정요인 .....	55

V. 논의 .....	61
1. 타자 .....	61
2. 투수 .....	63
3. 학문적·실무적 시사점 .....	64
VI. 결론 및 제언 .....	66
1. 결론 .....	66
2. 제언 .....	68
참고문헌 .....	70
Abstract.....	73



## 표 목 차

[표 II-1] 야구 경기 기록 표기 설명 .....	9
[표 III-1] 연구 변수 .....	25
[표 IV-1] 2012~2016시즌 KBO 타자 세이버메트릭스 지수 상관계수 .....	30
[표 IV-2] 2012~2016시즌 KBO 타자 주성분분석 결과 .....	31
[표 IV-3] 2012~2016시즌 KBO 타자 주성분분석 요인적재치 .....	31
[표 IV-4] 2012~2016시즌 KBO 타자 연봉 결정요인 다중회귀분석 결과 .....	33
[표 IV-5] 2004~2006시즌 KBO 타자 주성분분석 결과 .....	36
[표 IV-6] 2004~2006시즌 KBO 타자 주성분분석 요인적재치 .....	36
[표 IV-7] 2004~2006시즌 KBO 타자 연봉 결정요인 다중회귀분석 결과 .....	37

[표 IV-8] 2014~2016시즌 MLB 타자 주성분분석 결과.....	40
[표 IV-9] 2014~2016시즌 MLB 타자 주성분분석 요인적재치 .....	40
[표 IV-10] 2014~2016시즌 MLB 타자 연봉 결정요인 다중회귀분석 결과 .....	42
[표 IV-11] 2012~2016시즌 KBO 투수 세이버메트릭스 지수 상관계수 .....	46
[표 IV-12] 2012~2016시즌 KBO 투수 주성분분석 결과 .....	47
[표 IV-13] 2012~2016시즌 KBO 투수 주성분분석 요인적재치 .....	47
[표 IV-14] 2012~2016시즌 KBO 투수 연봉 결정요인 다중회귀분석 결과 .....	49
[표 IV-15] 2004~2006시즌 KBO 투수 주성분분석 결과 .....	52
[표 IV-16] 2004~2006시즌 KBO 투수 주성분분석 요인적재치 .....	52
[표 IV-17] 2004~2006시즌 KBO 투수 연봉 결정요인 다중회귀분석 결과 .....	54

[표 IV-18] 2014~2016시즌 MLB 투수 주성분분석 결과.....	57
[표 IV-19] 2014~2016시즌 MLB 투수 주성분분석 요인적재치 .....	57
[표 IV-20] 2014~2016시즌 MLB 투수 연봉 결정요인 다중회귀분석 결과 .....	59

## 그 립 목 차

[그림 Ⅲ-1] 최OO의 통산 기록.....	23
[그림 Ⅳ-1] 2012~2016년 KBO 타자 세이버메트릭스 지수 상관관계 산점도.....	30
[그림 Ⅳ-2] 2012~2016년 KBO 투수 세이버메트릭스 지수 상관관계 산점도.....	46

# I. 서론

## 1. 연구의 필요성

한국프로야구리그는 1982년 6개 구단으로 시작하여 비약적인 발전을 이루었다. 구단 수는 10개로 증가하였고, 한국 프로스포츠 종목 중 최다 관중수 기록, 중계권료의 상승, 방송시스템의 발전 및 FA제도의 시행 등 다양한 요소에서 발전이 있었다(최상진 & 박노준, 2015). 또한 2008 베이징 올림픽 야구 금메달, 2009 World Baseball Classic 준우승, 2010 광저우 아시안게임 야구 금메달 등 각종 국제대회에서의 우수한 성적은 한국프로야구 선수들의 경기력이 과거와 비교하여 많은 발전이 있었음을 시사한다. 이러한 국제대회에서의 성과는 2000년대 중반 침체되었던 한국프로야구의 인기를 회귀시켰다(이장택, 2015; 배재영 외, 2012). 그 결과 2016년 KBO리그는 역대 최고 관중 수인 830만 관중을 기록하였다.

1980년대 초반 프로스포츠의 도입은 선수들의 경기력뿐만 아니라 시장의 양적, 질적 성장을 도모하였다. 스포츠 시장 규모의 확대와 더불어 KBO리그 선수들의 경기력 수준이 높아지면서 자연스럽게 선수들의 가치 또한 지속적으로 상승하였다. 한국프로야구가 개막한 1982년 전체 선수단의 평균 연봉은 1,200만원을 기록한 반면, 2017년 현재 1군 등록선수 전체 평균 연봉은 2억원을 넘어섰다(한국야구위원회, 2017). 성과에 따른 연봉의

상승은 선수들에게 강한 동기부여가 되어 긍정적인 영향을 미치기도 하지만(Richard & David, 2003) 현재 KBO리그는 타당한 선수의 가치 평가 없이 지나치게 선수들의 몸 값이 상승했다는 지적을 받기도 한다(배정섭 외, 2015). 또한 이러한 상황 속에서 각자의 이익을 추구해야하는 구단과 선수는 연봉 협상 과정에서 발생하는 의견차이 따른 갈등을 경험하기도 한다(명왕성 외, 2016). 갈등의 원인은 본질적으로 선수들의 가치를 평가하고 연봉을 책정하는 데에 있어 객관적이고 체계적인 기준이 미흡하기 때문이다(신문선, 2002; 최상진 & 박노준, 2015).

객관적이고 체계적인 책정 기준이 확립되지 않은 상황에서의 연봉 협상은 리그내 연봉격차 심화를 야기할 수 있다. 실제로 2017년 KBO 리그의 구단 별 연봉 상위 27명(외국인선수 제외)의 평균 연봉은 2억 3,987만원인 반면, 상위 270명을 제외한 나머지 260명의 2017년 평균 연봉은 3,390만원에 불과하다(한국야구위원회, 2017). 이렇듯 KBO리그의 다차원적인 발전에도 불구하고 선수들의 가치를 평가하는 객관적인 기준은 미흡한 실정이다. 이와 같은 상황에서의 급격한 연봉의 상승은 다음과 같은 부작용을 야기할 수 있다.

첫째, KBO리그 선수들 사이에서의 연봉 격차 심화는 상대적으로 낮은 연봉을 받는 선수들에게 박탈감을 초래할 수 있다(신문선, 2002). 팀내 연봉 격차 심화로 인하여 발생하는 갈등 요소들의 누적은 팀워크를 약화시킬 뿐 아니라 선수단의 불화를 조성하여 팀 전체 경기력에 악영향을 끼칠 수 있다. (박성희 & 강준호, 2005; 김정우 & 김기민, 2011). 특히 MLB 팀의

임금격차와 승률에 관한 선행연구 대부분에서, 임금불평등이 클수록 팀 경기력에 부정적인 영향을 미치는 것을 확인할 수 있었다. (Richards & Guell, 1998; Depken, 2000; Todd Jewell & Molina, 2004). 이렇듯 연봉 격차의 심화는 선수 개인의 문제에서 더 나아가 경기력 하락에 따른 구단 경쟁력에도 부정적인 영향을 미칠 수 있다.

둘째, 선수들의 과도한 연봉은 구단 경영에 재정적 부담이 될 수 있다. 재정적 부담은 신인 선수의 육성, 팬들에 대한 서비스 차원의 보상 등 구단 차원에서 필요한 소비 및 투자를 저해할 수 있다. 이에 대하여 이장영 (2009)과 신문선(2002)은 KBO리그 선수들의 연봉이 과거와 비교하여 기하급수적으로 증가하여 대부분의 KBO 구단들은 구단경영에 어려움을 겪는다고 지적하였다.

다양한 선행연구들에서 알 수 있듯이 KBO리그의 급격한 연봉 인플레이션(inflation)은 리그 발전에 부정적인 영향을 미칠 수 있다. 이에 따라 프로스포츠 시장에서 선수들의 가치라고 여길 수 있는 연봉이 어떠한 요소에 의하여 결정되는지에 대한 탐색이 필요하다. 또한 현재 KBO리그 선수들의 연봉 결정요인 탐색을 통하여, 연봉 시장 전체에 대한 진단이 이루어져야 한다. 하지만 이러한 필요성에도 불구하고 KBO리그 연봉에 관한 선행연구들을 살펴보면 선수의 가치를 경기력만으로 한정하여 연봉과의 선형관계를 살펴본 연구가 대부분이다(이만규, 2006; 이장영 & 강효민, 2001; 오광모 & 이장택, 2003; 승희배 & 강기훈 2012). 이와 같은 선행연구들에서는 한국프로야구리그 선수들의 연봉이 경기력 요인만으로

결정된다고 단정하기 어렵다는 점을 언급하면서 선수의 인지도, 관중 동원 능력, 나이 등 경기력 이외의 요인들이 연봉 책정에 미치는 영향에 대한 연구가 필요성을 강조하였다.

이에 따라서 본 연구에서는 통계적 접근을 통해 KBO 선수들의 가치 평가는 어떠한 요소에 의해 이루어지고 있는지에 대한 분석을 실시하였다. 연봉 책정시 선수의 가치 평가에 대한 뚜렷한 기준이 없다면 불필요한 연봉의 인플레이션(Inflation)을 야기할 수 있기에, 연봉이 어떠한 요인에 의하여, 얼마만큼 설명되는지에 대한 분석이 필요하다.

본 연구는 타당하고 신뢰로운 프로야구 선수의 가치 평가를 위하여 한국프로야구리그 연봉 결정요인에 대하여 분석하고 진단하여 리그 특성에 맞는 연봉 책정 지표 개발을 위한 기초자료를 제시하고자 하였다. 이를위해 2012~2016시즌 KBO리그 선수들을 대상으로 경기력, 올스타 팬 투표 득표수로 설정한 인지도, 포지션, 선수경력, 나이, 올스타 출전 횟수, 좌타우타 구분, 좌투우투 구분, FA자격 구분, 투수의 키, 타자의 몸무게, 팀 구분에 대한 자료를 독립변수로 설정하여 연봉과의 관계를 살펴보았다. 또한 최근 5년의 기록만으로는 KBO리그의 연봉 결정요인에 대한 예측안정성을 확인하기 어렵기에, 2004~2006시즌 KBO리그 선수들을 대상으로 같은 절차의 분석을 실시하여 시기에 따른 연봉 결정요인에 대한 변화를 탐색하였다. 이와 같은 비교 분석은 시기에 따른 한국프로야구리그 연봉 결정요인의 변동을 설명해 줄 수 있을것이다.

또한 KBO리그 자료만을 대상으로 수행한 분석 결과로 현재 KBO리그의



연봉 책정이 타당한지에 대한 판단은 한계가 있으므로, KBO리그보다 역사가 깊고, 시장의 규모가 큰 MLB와의 비교 분석을 실시하였다. MLB와 KBO리그 연봉 연봉 결정요인 비교 분석은 리그 특성에 따른 연봉 결정요인의 차이를 탐색하여, 한국프로야구리그의 연봉 시장에 대하여 진단해 보았다는 점에서 그 의의를 찾을 수 있을 것이다.

## 2. 연구목적

한국프로야구 선수들의 경기력과 연봉의 관계를 살펴본 연구는 다양하게 진행되어왔다. 하지만 프로스포츠의 특성상 경기력만으로 연봉을 책정하기에는 무리가 있다(박승현, 2008; 승희배 & 강기훈, 2012; 박성배 외, 2016). 그럼에도 불구하고 경기력과 경기력 이외의 요소들을 함께 독립변수로 설정하여 연봉과의 관계를 살펴본 연구는 미흡한 실정이다. 따라서 본 연구에서는 KBO 선수들의 경기력을 포함한 다양한 연봉 결정요인들을 탐색하고, 분석 결과를 토대로 한국프로야구리그에 알맞은 연봉 책정 지표 개발을 위한 기초자료를 제시하는 데에 목적이 있다. 본 연구는 다음과 같은 세부 목적을 가지고 있다.

- 1) KBO리그와 MLB 선수들의 연봉 결정요인을 비교 분석한다.
- 2) 2012~2016 KBO리그와 2004~2006 KBO리그 연봉 결정요인의 변화 추이를 비교 분석하여 연봉 결정요인들의 안정성을 검정한다.
- 3) KBO리그와 MLB 선수들의 연봉 결정요인의 회귀계수를 비교하여 각 리그 연봉 책정의 특성을 확인한다.
- 4) 연구 결과에 따른 연봉 결정요인을 기반으로, 한국프로야구리그 선수들의 타당한 연봉 책정을 위한 기초자료를 제시한다.

### 3. 용어의 정의

#### 1) 세이버메트릭스 지수 (Sabermetrics Index)

세이버메트릭스는 야구 경기에서 발생하는 기록들을 통계적, 수학적으로 분석하는 기법을 말한다. 본 연구에서는 야구 경기의 단순기록을 바탕으로 가공된 경기력을 세이버메트릭스 지수로 정의하였다. 본 연구에 포함된 세이버메트릭스 지수들은 타자의 경우 wOBA, WAR, RC27, XR27, BB/K, GPA, SLG, OBP이고, 투수의 경우 WHIP, OBP, SLG, FIP, ERA, WAR이다. 각 지수들에 대한 정의는 이론적 배경에 제시되어있다.

#### 2) 경기력(Performance)

본 연구에서는 주성분분석을 통하여 선수의 경기력을 나타내는 세이버메트릭스 지수들을 하나 혹은 둘의 벡터로 축약하여 사용하였다.

#### 3) 인기도(Popularity)

본 연구에서는 한국야구위원회 홈페이지에서 제공하는 올스타팬투표 결과를 바탕으로 매 시즌 최고 득표자를 100점으로 계산하여, 선수의 인기를 나타내는 변수로 설정하여 사용하였다.

## Ⅱ. 이론적 배경

### 1. 야구에서의 경기력 지표

야구경기에서 발생하는 수많은 단순 경기기록들은 점차 누적되어 선수 평가에 있어 통계적 가치를 갖게된다(이장택, 2014). 하지만 단순 경기 기록은 선수의 경기력에 대하여 한정적이고 단편적으로만 나타내주기 때문에, 현대 야구에서는 세이버메트릭스를 활용하여 선수들의 경기력을 평가하고 있다. 세이버메트릭스란 선수의 누적된 경기기록을 바탕으로 통계적으로 분석하는 방법론을 통칭한다(Baumer & Zimbalist, 2013). 이러한 세이버메트릭스를 활용한 분석은 구단과 선수의 연봉 협상 시에도 선수의 경기력을 평가함에 있어 타당하고 신뢰로운 자료를 도출해 낼 수 있다(양도업 외, 2015). 이와 같은 선수의 경기력을 나타내는 타당한 지표는 선수 간 발생하는 임금불균형을 해소할 수 있을 것이다. 따라서 현대 프로야구시장에서 세이버메트릭스의 중요성은 더욱 커지고 있다.

세이버메트릭스는 1971년 빌 제임스(Bill James)를 비롯한 학자들에 의해 창립된 미국야구연구회(SABR: Society for American Baseball Research)에서 만들어진 단어이다. 세이버메트릭스는 야구 경기에서 발생하는 기록들을 활용하여 통계적, 수학적으로 분석하는 기법을 의미한다. James (1987)는 세이버메트릭스를 ‘야구에 대한 객관적 지식을 찾고자 하는 연구’

라고 정의하였고 그 목적은 선수의 가치와 경기력을 정량화하는 것이라고 하였다.

〈표 II -1〉은 세이버메트릭스 계산식에 대한 설명에 앞서 단순 기록에 대한 표기에 대하여 정리한 표이다.

표기	기록 설명
H	‘1루타+2루타+3루타+홈런’의 수
BB	Base on balls의 약자로 볼넷(4구)의 수
IBB	고의4구의 수
IP	투수의 투구 이닝
SO	삼진 아웃 수
AB	타자가 타석에 들어선 수
SF	희생플라이 수
SC	희생 번트 수
CS	도루 실패 수
SB	도루 성공 수
GIDP	땅볼로 인한 병살타의 수
HBP	몸에 맞은 볼의 수
RBI	타점의 수

〈표 II -1〉 야구 경기 기록 표기 설명

나열한 단순 경기 기록을 활용한 투수와 타자의 경기력을 나타내는 세이버메트릭스 지수에 대한 구체적인 설명과 계산방식은 다음과 같다.

## 1) 투수의 세이버메트릭스(Sabermetrics) 지수

### (1) WHIP

Walks plus Hit Divided by Innings Pitched의 약자로, 이닝당 출루허용율을 의미한다. 계산 식은 다음과 같다.

$$WHIP = (BB+H) / IP$$

피안타와 볼넷 허용 개수를 이닝으로 나눈 수치이다. 단순한 계산으로 쉽게 투수의 성적을 쉽게 이해하고 평가할 수 있기에 최근 야구 중계 방송에 자주 나타나는 기록이다. 하지만 몸에 맞는 볼과 장타를 평가 요소에서 배제했다는 비판 또한 따른다 (이장택, 2017).

## (2) K/9

9이닝 동안의 탈삼진 개수를 의미한다. 9이닝을 기준으로 하였기에 선발투수의 삼진 능력을 평가하는 기준이 된다. 계산식은 다음과 같다.

$$K/9 = 9(K) / IP$$

9를 곱하는 이유는 9이닝을 던졌을 때를 기준으로 보기 위함이다.

## (3) BB/9

9이닝 동안 허용한 볼넷의 개수를 의미한다. 계산식은 다음과 같다.

$$BB/9 = 9(BB) / IP$$

K/9과 마찬가지로 9이닝을 기준으로 볼넷의 개수를 보기 위한 수치이고, 투수의 제구력을 평가하는 기준으로 쓰인다.

## (4) BABIP

인플레이로 이어진 타구에 대한 타율을 계산하는 수치이다. 계산식은 다음과 같다.

$$BABIP = (H-HR) / (AB-K-HR+SF)$$

홈런, 삼진을 제외한 인플레이 타구는 수비수의 능력이 투수 평가에 영향을 미치기에 BABIP가 야구 통계학자들에 의해 고안되었다. 투수의 경우 해당 수치가 리그 평균보다 높으면 수비의 도움을 받지 못한 것으로 판단할 수 있다.

#### (5) FIP

Fielding Independent Pitching의 약자로 BABIP와 달리 투수가 통제가능한 영역만으로 투수를 평가하는 지수이다. 계산식은 다음과 같다.

$$FIP = \frac{13(HR)+3BB-2K}{IP} + C$$

홈런, 삼진을 제외한 인플레이 타구는 투수기록에 수비능력이 반영된다. FIP는 투수가 통제가능한 영역인 삼진, 볼넷, 홈런기록만으로 투수의 능력을 나타낸다. 팀의 수비능력이 반영되는 지수인 ERA와 같은 형태로 나타내기, 투수의 독립적인 경기력을 ERA와 비교하여 쉽게 살펴볼 수 있다는 장점이 있다.(이장택, 2017). C는 Constant의 약자로 리그 ERA와 비슷한 값을 갖도록 더해주는 상수이다. 일반적으로 3.20 정도의 값을 더해준다.

## 2) 타자의 세이버메트릭스(Sabermetrics) 지수

### (1) OPS

On base percentage Plus Slugging percentage의 약자로 세이버메트릭스 지수중에서 가장 보편화 된 타자 평가 지수이다. 계산식은 다음과 같다.

$$OPS = OBP + SLG$$

$$OBP = \frac{H+BB}{AB+BB+SF}, \quad SLG = \frac{\{1B+(2B*2)+(3B*3)+(HR*4)\}}{AB}$$

출루율과 장타율의 합으로 계산할 수 있는 간편한 공식이기에 타자를 평가함에 있어 간단하면서도 매우 효과적인 지수라고 할 수 있다. 장타율과 비교하여 출루율이 저평가 되는 단점을 지니고 있지만, 타자의 공격능력을 종합적으로 평가해주기에 프로구단 및 방송 매체에서 많이 사용하는 지수이다.

### (2) GPA

Gross Production Average의 약자로 타자의 출루율에 1.8의 가중치를 곱한 값과 장타율을 더한 값에 나누기 4를 한 값을 의미한다. 4를 나누어주는 이유는 해당 값을 타율의 범위와 비슷하게 나타내기 위함이다. 해당 지수는 앞선 OPS지수에서 출루율이 저평가됨을 보완하고자 고안된 지수이다. 계산식은 다음과 같다.



$$GPA = \frac{1.8*OBP+SLG}{4}$$

$$OBP = \frac{H+BB}{AB+BB+SF}, \quad SLG = \frac{\{1B+(2B*2)+(3B*3)+(HR*4)\}}{AB}$$

### (3) BB/K

타자가 얻은 볼넷의 수를 삼진 수로 나눈 값으로, 타자의 출루능력을 평가하는 통계량이다. 계산식은 다음과 같다.

$$BB/K = \text{Base on balls} / \text{Strike Out}$$

### (4) XR 27

eXtrapolated Runs 27의 약자로 선수 혼자 만들어낸 안타, 홈런, 번트등과 같은 기본 통계량을 이용하여 선수의 득점생산성을 측정하기 위한 세이버메트릭스 통계량이다. 계산 식은 다음과 같다.

$$XR27 = \frac{XR}{PA-H+SH+SF+CS+GIDP} \times 27$$

$$XR = 0.50(1B)+0.72(2B)+1.04(3B)+1.44HR+0.34(HBP+TBB-IBB)+0.25IBB \\ +0.18SB-0.32CS-0.09(AB-H-K)-0.098K-0.37GIDP+0.37SF+0.04SH$$

(5) RC 27

Runs Created per 27 outs 의 약어로 27개의 아웃동안 해당 타자가 1~9번 타순에 모두 섰을 때, 몇 점을 득점할 것인가를 평균 수치화 한 것이다. 계산식은 다음과 같다.

$$RC27 = \frac{RC}{(AB-H+SH+SF+CS+GIDP)} \times 27$$

$$RC = \frac{A \times B}{C}$$

$$A = H + BB + HBP - CS - GIDP$$

$$B = TB + 0.26(BB-IBB+HBP) + 0.52(SH+SF+SB)$$

$$C = AB + BB + HBP + SH + SF$$

## 2. 프로스포츠 연봉 결정요인 분석 방법

연봉 결정요인 분석이란 한 선수의 현재 연봉이 어떠한 요소에 의하여 결정되는지에 대한 분석이다. 선수들의 연봉은 한 가지 요인으로만 설명되는 것이 아니라 경기고과, 구단의 인사 평가 등 다양한 요인들에 의해 책정된다(승희배 & 강기훈, 2012; 신문선, 2002). 일반적으로 KBO 리그에서는 선수경력, 나이와 경험, 포지션, 경기력 및 팀 승리 기여도, 팬들에게 미치는 영향력 등에 의해서 선수의 가치가 평가된다(박성배 외, 2016). 하지만 대부분의 선행 연구들은 연봉과 경기력의 관계만을 다루고 있고, 경기력 이외를 요소를 포함하여 연봉과의 선행관계를 살펴본 연구는 미흡한 실정이다.

프로스포츠 연봉과 관련한 선행 연구들을 살펴본 결과 연봉이 어떠한 요소에 의해 결정되는지 파악하기 위하여 주성분분석, 다중회귀분석, 데이터마이닝 기법을 사용한 연구가 주를 이루었다. 본 연구에서는 주성분분석과 다중회귀분석을 중심으로 연구를 진행할 예정이므로 두 가지 분석을 사용한 선행 연구를 살펴보았다.

### 1) 주성분분석을 활용한 연봉 결정요인 탐색

주성분분석은 2개 이상의 반응변수로 얻어진 다변량 자료를 대상으로 다차원 변수들의 차원 축소, 요약과 더불어 서로 상관성이 있는 반응변수들 간의 복잡한 구조를 규명하는 분석 방법이다(김기영, 1989). 분석의 목적은

주어진 다변량 자료를 단순화하고 정보의 손실을 최소화 하면서 차원축소를 통하여 변수의 관계를 규명하는 것이다(이장택, 2014).

주성분분석에 대한 개념을 가장 처음 제안한 연구자는 Pearson (1901)으로, 직교최소제곱(orthogonal least squares)의 개념을 기반으로 다차원 공간상에 놓이는 점들을 설명할 수 있는 최적의 직선이나 평면을 찾기위한 기하학적 최적화(optimization)의 문제로 접근하였다. 그 후 Hotelling(1933)은 원자료들의 분산을 최적으로 설명하는 성분을 유도하여 산출하였고, 산출된 성분들을 ‘주성분(principal component)’이라고 하였으며, 이에 기초한 분석을 ‘주성분분석’이라고 명명하였다. 이와 같은 주성분분석은 공분산행렬을 이용한 방법과 상관행렬을 이용한 방법으로 나눌 수 있는데, 보편적으로 주성분분석에 투입된 변수들의 단위가 다르거나 각 변수들의 분산의 차이가 클 경우 공분산행렬보다는 상관행렬을 이용한 주성분분석을 실시한다(김재희, 2011). 산출된 각 주성분의 방향은 공분산행렬 혹은 상관행렬의 고유벡터(eigenvector)로 주어지며, 각 주성분의 방향에 따른 분산은 고유값(eigenvalue)으로 주어진다. 예를들어 특정 변수의 첫번째 주성분 점수는 첫번째 고유벡터(eigenvector)와 변수의 데이터를 곱한 값으로 나타난다. 이와 같은 주성분분석을 통하여 산출된 주성분은 투입된 표본변수에서 나타내는 변동 중 가능한 많은 변동을 설명할 수 있는 요인들에 대한 최적의 선형결합을 의미한다(김기영, 1989).

승희배와 강기훈(2012)의 연구에서는 경기력을 나타내는 세이버메트릭스 지수들을 주성분분석을 통하여 주성분에 대응되는 고유의 벡터 값을 산출한

뒤 해당 값을 활용하여 연봉과의 관계를 살펴보았다. 이제영 & 김현규(2016)의 연구에서는 주성분분석을 활용하여 WAR을 대체 할 새로운 타자력지수를 제안하였고, 이장택(2014), 홍종선 외(2016) 또한 주성분분석을 활용하여 경기력을 하나의 요인으로 설정하여 연구를 진행하였다.

본 연구에서 역시 다양한 세이버메트릭스 지수들에 대하여 차원 축소를 위하여 상관관계수행렬을 이용한 주성분분석을 실시하였고, 산출된 주성분을 프로야구 선수의 연봉을 종속변수로 하는 다중회귀모형의 독립변수로 설정하여 분석을 진행하였다.

## 2) 선형회귀분석을 활용한 연봉 결정요인 탐색

선형회귀분석은 예측과 관련한 분석에 있어서 가장 보편화된 기법으로, 모형을 통해 도출한 회귀방정식은 상관관계를 기초로 하여 독립변수로부터 종속변수를 설명한다. 아울러 t값 및 유의확률을 통해 독립변수의 설명 방향성과 유의성을 분석한다(Linoff & Berry, 2011). 회귀분석시 종속변수를 예측 혹은 설명해주는 독립변수가 한 개일 경우에는 단순회귀분석, 독립변수가 두 개 이상일 경우 다중회귀분석이라 한다(Cohen et al, 2013).

프로스포츠 선수의 연봉 결정요인을 탐색한 선행 연구들은 대부분 회귀분석을 사용하여 수행하였다. 이장영과 최병목(2002)은 단순회귀 분석과 다중회귀분석을 활용하여 프로농구 선수의 경기력이 연봉에 미치는 영향에 대하여 분석하였다. 해당 연구의 분석 방법에서 단순회귀분석과

다중회귀분석을 모두 사용한 이유는 다양한 독립변수들을 단순히 각각 연봉과의 상관을 파악했을 때와 모두 함께 변수로 포함되어 연봉을 설명하려고 할 때와 어떤 결과의 차이가 나타나는지를 파악하기 위함으로 판단된다. 박성배 외(2016)는 종속변수를 연봉으로 설정하여 다양한 독립변수와의 선형관계를 살펴보았다. 투입된 독립변수는 FA유무, RC/27, SecA, 연령이다. 해당 연구에서 FA유무를 독립변수로 설정한 이유는 자유계약선수의 구분 없이 일괄적으로 선수 평가 요인을 적용하는 것은 타자들의 가치를 평가함에 있어 객관성을 떨어뜨릴 수 있는 잠재적 요인으로 작용할 수 있기 때문이라고 설명하고있다.

이렇듯 연봉 결정요인 탐색과 관련한 선행 연구들은 대부분 회귀분석을 통하여 연봉과 투입된 독립변수의 관계를 살펴보고 있다. 하지만 경기력만으로 선수의 연봉을 예측하는 것은 한계가 따르기에, 본 연구에서는 다중회귀분석을 위하여 설정한 다양한 변수 중 ‘Salary’ 변수를 종속변수로 사용하였고, ‘Salary’이외의 모든 변수를 독립변수로 설정하여 분석을 실시하였다.

### 3. 프로스포츠 연봉 결정요인 관련 선행 연구

프로스포츠 선수들의 연봉 결정요인을 주제로 진행한 연구를 살펴보면 다음과 같다. 김종환 외(2013)의 연구에서는 ‘팀 성적에 대한 선수의 공헌도와 팀 수입에 미친 효과’로 정의한 선수의 한계생산(Scully, 1974)이 선수의 연봉을 결정한다고 하였다. 선수개인의 역량이 투입된 팀 성적에 따라 팀 수입이 결정된다는 전제하에 팀 성적에 기여한 개인의 공헌도에 따라 선수의 연봉을 결정해야 한다고 하였다. 신문선(2004)은 한국 프로스포츠를 대표하는 프로축구와 프로야구의 연봉산정 방법에 관하여 정량화 정도와 객관적인 평가방법에 대한 요인별 비교를 시행하였다. 특히 프로야구 구단의 경우 선수의 연봉을 산정하는데 있어 각 구단이 자체적으로 책정하는 경기고과 점수를 가장 많이 활용하고 있다고 언급하며, 선수의 연봉 산정에 대하여 보다 과학적이고 체계적인 연봉산정 기준의 제시 및 합리적인 연봉산정모델 개발이 필요성을 강조하였다. 또한, 한필수(2006)의 연구에서는 기존 프로농구 선수들의 경기실적을 토대로 연봉액을 산정할 수 있는 모형에 스포츠의 경기를 질적으로 평가할 수 있도록 생성, 확대, 개발된 경기요인을 이용하여 모형개발을 시도하였다.

연봉 결정요인을 중심으로 한 연구는 국내 뿐 아니라 국외에서도 활발히 진행되고 있는 분야이다. Leeds, Sakata와 Von Allmen(2012)은 일본프로야구(NPB: Nippon Professional Baseball)의 선수 연봉 결정요인을

탐색하였는데, 해당 연구에서는 타자의 포지션에 따른 연봉 결정요인의 여러 특성을 분석하였다. 또한 Jane(2010)은 MLB(Major League Baseball)의 데이터를 이용하여 팀의 연봉 구조 및 해당 경기력과 관계를 규명하며 경기력을 향상시키는 방법으로 급여를 확대하는 것보다 급여를 줄이는 것임을 새롭게 제시하기도 하였다. Tao, Chuang과 Lin(2016)의 연구에서는 1985-2013년 MLB 데이터를 기반, 토너먼트 이론(tournament theory)과 팀 응집력 가설(team-cohesiveness hypothesis)을 중심으로 연봉과 팀 경기력 간 관계를 설명하였다.

반면 경기력 이외의 요소를 투입하여 프로 선수들의 연봉 결정요인을 심도있게 다룬 연구는 매우 부족한 실정이다. 이장영과 최병목(2002)의 연구에서는 한국 프로농구선수의 경기능력과 연봉의 관계를 이해하는 연구를 진행하며 경기력 못지 않게 선수 개인의 인기 역시 연봉산정에 영향을 미치는 변수로 여겨진다고 하였다. 하지만 인기를 나타낼만한 적절한 자료를 얻지 못하여 변수로 추가하지 못하여 쉽게 자료를 얻을 수 있는 변수들로 연구를 제한하였다. 박승현(2008)의 연구에서도 이와 유사한 제한점이 나타나는데, 한국 프로야구 타자들의 고액연봉 결정에 경기력의 어떤 요인이 가장 커다란 영향을 미치는가를 분석하였지만 고액연봉이 경기력 요인만으로 결정된다고 단정하기는 어렵다고 제시하였다.

많은 선행연구에서 연봉 결정요인으로 경기력을 공통적으로 제시하고 있으며 경기력 이외의 요소가 중요하다고 제시하고 있음에도 불구하고, 해당 요소가 어떻게 연봉 결정요인으로 형성되며 연봉에 영향을 미치는지



검증할 수 있는지에 대한 연구는 매우 미흡하다. 특히 리그를 대표하는 선수의 관중동원능력이나 올스타 선정, 올스타 팬 투표수 등의 요인은 선수의 연봉 책정에 상당한 영향을 미칠 수 있는 요인이다. 하지만 경기력과 함께 독립변수로 설정하여 진행된 연구는 매우 미흡한 실정이다.

다시 말해, 프로스포츠 선수들의 연봉 결정요인을 중심으로 한 연구에 있어 경기력과 경기력 이외의 요소를 동시에 분석하여 어떠한 요소가 연봉에 영향을 미치는지에 대한 통합적인 연구가 필요하다.

### Ⅲ. 연구 방법

본 연구의 목적은 KBO리그와 MLB선수들의 연봉 결정요인을 파악하고, KBO리그에 알맞는 연봉 책정을 위한 기초 자료를 제시하는 것이다.

연구 수행을 위한 KBO리그 경기기록은 한국야구위원회 홈페이지([www.koreabaseball.com](http://www.koreabaseball.com))를, 연봉 자료는 KBO 가이드 북과 한국프로야구 리그 경기 기록 및 통계를 전문적으로 다루는 Statiz 홈페이지([www.statiz.co.kr](http://www.statiz.co.kr))에서 수집하였다. 또한 MLB의 자료는 공식 기록 집계 홈페이지인 MLB 공식 홈페이지([www.MLB.com](http://www.MLB.com))와 MLB 전문 통계 기록 사이트인 베이스볼레퍼런스 홈페이지([www.baseball-reference.com](http://www.baseball-reference.com))에서 수집하였다.

본 연구에서는 상관행렬을 이용한 주성분분석을 통하여 경기력을 나타내는 다양한 세이버메트릭스 지수들을 타자 분석모형에서는 하나의 변수로, 투수 분석모형에서는 두 개 변수로 설정하였다. 다음 단계로 다중회귀분석을 위하여 선수의 경기력, 포지션, 인기, 경력, 나이, 좌투우투 구분(투수), 좌타우타 구분, 투수의 키, FA자격 구분, 올스타 출전 횟수, 올스타 팬 투표 득표수(KBO타자)의 변수들을 독립변수로 설정하고, 종속변수는 연봉으로 설정하여 연봉과 다양한 독립변수들과의 선형관계를 살펴보았다.

본 연구에 사용된 방법과 절차의 세부적인 내용은 다음과 같다.

## 1. 연구 대상

본 연구에서는 2012~2016시즌, 2004~2006시즌 KBO리그 출전 선수와 2014~2016시즌 MLB 출전 선수를 대상으로 분석을 진행하였다. 타자의 경우 규정타석 수를 채운 타자, 투수의 경우 규정 이닝을 충족시킨 선수로 설정하였다. 규정타석과, 규정이닝 충족에 대한 기준 설정은 KBO 자체에서 규정타석과 규정이닝을 설정하여 선수 평가 기준으로 사용하고 있기에 이를 반영하여 대상을 설정하였다.

〈그림 Ⅲ-1〉은 KBO 공식홈페이지(www.koreabaseball.com)에서 발췌한 KBO 타자 최OO 선수의 공식 통산 기록이다. 〈그림 Ⅲ-1〉에서 볼 수 있듯이 공식홈페이지에서는 경기에 대한 단순 기록 및 세이버메트릭스 지수, 포지션, 좌타/우타 구분, 연봉, 데뷔년도 등 연구에 필요한 다양한 자료를 수집할 수 있다.

타자

투수

기록용어

기본기록

통산기록

일자별기록

경기별기록

상황별기록

KBO 정규시즌

KBO 정규시즌

연도	팀명	AVG	G	AB	R	H	2B	3B	HR	TB	RBI	SB	CS	BB	HBP	SO	GDP	SLG	OBP	E
2002	삼성	0.400	4	5	0	2	2	0	0	4	0	0	0	0	0	0	0	0.800	0.400	0
2004	삼성	0.000	2	2	0	0	0	0	0	0	0	0	0	0	0	1	0	0.000	0.000	0
2008	삼성	0.276	126	384	68	106	24	0	19	187	71	3	0	48	6	76	9	0.487	0.364	3
2009	삼성	0.284	113	415	70	118	24	0	23	211	83	1	2	53	9	60	12	0.508	0.375	3
2010	삼성	0.279	121	420	71	117	29	1	24	220	97	4	0	71	8	91	13	0.524	0.389	2
2011	삼성	0.340	133	480	80	163	37	3	30	296	118	4	3	76	5	88	8	0.617	0.427	1
2012	삼성	0.271	125	461	51	125	27	1	14	196	77	2	1	55	5	76	20	0.425	0.348	3
2013	삼성	0.305	128	511	80	156	28	0	29	271	98	2	1	47	7	91	5	0.530	0.366	1
2014	삼성	0.356	113	430	92	153	33	0	31	279	100	4	2	50	7	62	11	0.649	0.426	1
2015	삼성	0.318	144	547	94	174	33	1	33	308	123	2	5	73	9	101	13	0.563	0.402	2
2016	삼성	0.376	138	519	99	195	46	2	31	338	144	2	2	83	9	83	12	0.651	0.464	3
2017	KIA	0.300	7	20	6	6	2	1	1	13	5	0	1	5	1	3	0	0.650	0.444	0
통산		0.314	1154	4194	711	1315	285	9	235	2323	916	24	17	561	66	732	103	0.554	0.398	19

〈그림 Ⅲ-1〉 최OO의 통산 기록 (출처: 한국야구위원회 기록실)

## 2. 연구 도구

본 연구에 사용된 도구의 구체적 명칭은 다음과 같다.

자료저장 프로그램 : Microsoft Office Excel 2013

다중회귀분석 프로그램 : R Studio 1.0.136 - Mac OS X 10.6+

주성분분석 프로그램 : R Studio 1.0.136 - Mac OS X 10.6+

## 3. 자료의 구성

한국프로야구 연봉 관련 선행연구들은 대부분 경기력만으로 연봉과의 관계를 살펴보았다. 하지만 다수의 연구자들이 선수의 인기도, 선수의 경력, 포지션 등의 변수들을 함께 독립변수로 설정하여 연봉과의 관계를 살펴볼 필요성을 강조하고 있다(박승현, 2008; 승희배 & 강기훈, 2012; 박성배 외, 2016). 따라서 본 연구에서는 경기력 이외의 다양한 요소를 독립변수로 설정하였다. 본 연구에 사용된 변수들은 <표 Ⅲ-1> 같다.

변수명	변수에 대한 설명
Salary	선수의 연봉을 의미한다. 본 연구에서는 선수기록의 다음 시즌 연봉을 종속변수로 설정하여, 전 시즌 기록이 다음 시즌 연봉에 어떠한 영향을 미치는지 선형관계를 살펴보았다.
Performance	타자의 경기력을 의미한다. 본 연구에서는 주성분분석을 이용하여 하나의 벡터로 설정된 경기력 점수를 사용하여 독립변수로 설정하였다.
Performance1	투수의 다양한 경기기록을 주성분분석을 통하여 하나의 벡터로 변수화한 투수의 경기력을 의미한다.
Performance2	투수의 승리관련 기록들을 주성분분석을 통하여 하나의 벡터로 변수화한 투수의 승리관련 경기력을 의미한다.
Popularity	선수에 대한 팬들의 선호도 및 인기를 의미한다. 한국프로야구 리그 올스타 팬 투표의 득표수를 활용하여 선수의 인기를 나타내는 변수로 설정하였다.
Career	선수의 경력을 의미한다. 본 연구에서는 선수의 프로입단년도를 바탕으로 변수를 설정하였다.
Allstar	각 리그 선수들의 올스타전출전횟수/경력을 의미한다. 본 연구에서는 해당 변수를 독립변수로 설정하였다.
Position N	선수의 포지션을 의미한다. 본 연구에서는 해당 변수를 야구 수비 포지션 번호에 따른 더미변수로 변환하여 독립변수로 설정하였다. 외야수의 경우 좌익수, 중견수, 우익수의 구분 없이 일괄적으로 Position 7로 변환하여 더미변수화 하였다.
Age	선수의 출생년도에 따른 해당 시즌의 선수의 나이를 의미한다.
Right	투수의 경우 우완투수와 좌완투수의 구분을, 타자의 경우 우타자와 좌타자의 구분을 위하여 범주형 변수로 코딩하여 독립변수로 설정하였다.
Height	선수의 신장을 의미한다. 본 연구에서는 투수의 회귀분석에 독립변수로 설정하여 분석을 진행하였다.
FA	FA자격 선수의 구분을 의미한다. 본 연구에서는 FA구분에 따른 연봉과의 관계를 살펴보기 위하여 범주형 변수로 코딩하여 독립변수로 설정하였다.

〈표 Ⅲ-1〉 연구 변수

#### 4. 자료처리 방법

본 연구에 사용된 자료는 R Studio 1.0.136 - Mac OS X 10.6+를 이용하여 분석하였으며 그 구체적인 과정은 다음과 같다.

- 1) KBO리그 선수들의 경기력을 나타내는 다양한 세이버메트릭스 지수들을 상관행렬을 이용한 주성분분석을 통하여 타자의 경우 하나의 벡터(Vector)를, 투수의 경우 두 개의 벡터를 산출하여 변수화하였다.
- 2) 경기력, 인기도(올스타팬투표 득표수), 선수경력, 포지션, 키(투수), 몸무게(타자), 좌투우투 구분(투수), 좌타우타 구분(타자), FA자격 구분, 올스타 출전 횟수, 팀 구분 등 다양한 변수들을 독립변수로 설정하여 연봉과의 선형관계를 살펴보기 위한 다중회귀분석을 실시하였다.
- 3) 위와 같은 절차의 분석을 2004~2006년 KBO리그 선수들의 자료로 수행 한 뒤, 산출된 회귀계수 비교를 통하여 2017년 기준 최근 5년과 10년 전 KBO리그의 연봉 결정요인들의 변동 추이에 대하여 검증하였다.
- 4) 2014~2016년 MLB 선수들의 자료로 위와 같은 절차의 분석을 실시 한 뒤 산출된 회귀계수를 KBO리그와 비교 분석 하였다.
- 5) KBO와 MLB 선수들의 연봉 결정요인 비교 분석 결과를 토대로 한국프로야구 연봉시장에 대하여 진단하고, 산출된 회귀계수를 바탕으로 국내 프로야구에 알맞는 연봉 책정 지수 개발을 위한 기초 자료를 제시하였다.

## IV. 연구 결과

본 연구에서는 프로야구 선수들의 가치 평가 요인을 탐색하기 위하여 다중회귀분석(multiple regression analysis)을 통해 선수들의 연봉이 어떠한 요소에 의하여 결정되는지 분석을 실시 하였다.

선행연구들을 살펴보면, 대부분의 연구에서 어떠한 경기력 요소가 연봉에 영향을 미치는지에 대한 분석이 주를 이루었다. 하지만 이와 같은 접근은 프로스포츠 선수들의 연봉에 영향을 미칠 수 있는 경기력 이외의 요소에 대해서는 설명하지 못한다는 한계점을 갖는다. 따라서 본 연구에서는 주성분분석(Principal Component Analysis)을 통하여 다양한 세이버 메트릭스 지수를 타자의 경우 경기력을 하나의 변수로, 투수의 경우 경기력을 두 개의 변수로 설정 한 뒤 경기력 변수와 경기력 이외의 변수가 프로야구 선수의 연봉에 어떠한 영향을 미치는지 다중회귀분석을 통하여 살펴보았다. 경기력 이외의 변수는 선수경력(Career), 나이(Age), 인기도(Popularity), 올스타출전 횟수/ 경력(Allstar), FA자격 구분(FA), 좌타우타/좌투우투 구분(Right), 신장(Height), 몸무게(Weight), 포지션 구분(Position X), 팀 구분(Team X) 등으로 설정하였다.

투입된 독립변수는 타자와 투수, KBO리그와 MLB의 구분에 따라 차이를 보이는데 이는 각 포지션의 특성과 리그 특성을 반영하여 분석을 진행하였기에 나타난 것이다. 특히 타자의 경우에는 체격에 따라서 타구의

비거리가 달라질 수 있다고 판단하여 몸무게(Weight)를 독립변수로 설정하였고, 투수의 경우 신장에 따라서 투구한 공의 질이 달라질 수 있다고 판단하여 키(Height)를 독립변수로 설정하였다. 또한 메이저리그의 경우에는 선수들의 국적이 매우 다양한 반면 KBO리그는 용병선수가 한정적이기에 이를 독립변수로 설정하여 분석을 진행하였다.

시기에 따른 KBO의 연봉 결정요인에 대한 분석과 리그에 따른 연봉 결정요인을 분석하기 위하여 2012~2016시즌 KBO리그, 2004~2006시즌 KBO리그, 2014~2016시즌 MLB리그 선수들을 대상으로 타자 회귀모형 3개, 투수 회귀모형 3개, 총 여섯개의 다중회귀모형을 설정하여 연봉 결정 요인 분석을 진행하였다. 구체적인 분석 결과는 다음과 같다.



## 1. 타자의 연봉 결정요인 분석

### 1) 2012 ~ 2016년 KBO리그 타자의 연봉 결정요인

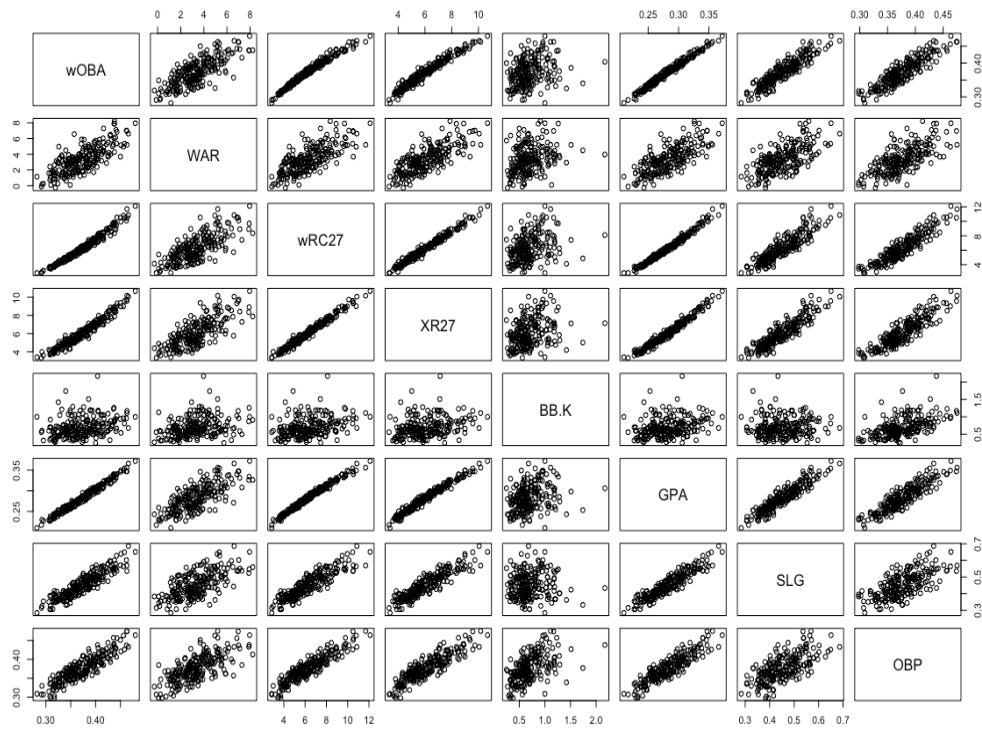
2012~2016년 KBO리그 타자의 연봉 결정요인을 분석하기 위하여 1차적으로 주성분분석을 실시하였다. 다음 단계에서는 다중회귀분석을 통하여 주성분분석에서 산출된 경기력 점수와 경기력 이외의 다양한 요인들을 독립변수로 설정하여 연봉과의 관계를 살펴보았다.

#### (1) 2012 ~ 2016년 KBO리그 타자 주성분분석 결과

본 분석에서는 8가지의 세이버메트릭스 지수를 축약하여 경기력 변수로 설정하기 위하여 주성분분석을 실시하였다. 해당 데이터가 주성분분석에 적합한 데이터인지를 확인하기 위하여 Bartlett의 구형성 검정을 수행한 결과 유의확률이 0.001보다 작게 나타났다. 이는 요인분석의 기본가정을 만족하는 결과이기에, 해당 데이터로 주성분분석을 실시하였다. <그림Ⅳ-1>은 주성분분석에 투입된 세이버메트릭스 지수 간의 상관관계를 산점도로 나타낸 것이다. 각 세이버메트릭스 지수 간의 상관계수는 <표Ⅳ-1>과 같다. <표Ⅳ-1>을 살펴보면, 상관계수의 최대값은 0.9898, 최저값은 0.0387로 8가지의 세이버메트릭스 지수 중 하나의 지수를 경기력이라고 설정하기 어려움을 알 수 있다.

이에 본 연구의 2012~2016시즌 KBO 타자 연봉 결정요인 회귀분석 모형에서는 주성분분석을 통하여 8개의 세이버메트릭스 지수의 최적조합을

구함으로써 타자의 경기력을 하나의 변수로 설정하였다. 주성분분석 결과는 <표Ⅳ-2>와 같다.



<그림Ⅳ-1> 2012~2016시즌 KBO 타자 세이버메트릭스 지수 상관관계 산점도

	wOBA	WAR	RC27	XR27	BB/K	GPA	SLG	OBP
wOBA	1.0000							
WAR	0.7611	1.0000						
RC27	0.9849	0.7367	1.0000					
XR27	0.9729	0.7585	0.9867	1.0000				
BB/K	0.2866	0.3001	0.3262	0.2929	1.0000			
GPA	0.9882	0.7353	0.9898	0.9834	0.2859	1.0000		
SLG	0.9145	0.6525	0.8952	0.9088	0.0387	0.9305	1.0000	
OBP	0.8923	0.6986	0.9190	0.8897	0.5299	0.8970	0.6727	1.0000

<표Ⅳ-1> 2012~2016시즌 KBO 타자 세이버메트릭스 지수 상관관계수

	주성분1	주성분2	주성분3	주성분4	주성분5	주성분6	주성분7	주성분8
Eigenvalue	6.334	1.058	0.394	0.164	0.029	0.015	0.006	0
% of var.	79.18	13.23	4.92	2.05	0.36	0.18	0.08	0.001
Cumulative % of var.	79.18	92.41	97.33	99.38	99.74	99.92	99.99	100

〈표Ⅳ-2〉 2012~2016시즌 KBO 타자 주성분분석 결과

	wOBA	WAR	RC27	XR27	BB/K	GPA	SLG	OBP
주성분 1	0.988	0.805	0.99	0.986	0.364	0.99	0.893	0.921
주성분 2	-0.083	0.047	-0.037	-0.076	0.913	-0.09	-0.365	0.259

〈표Ⅳ-3〉 2012~2016시즌 KBO 타자 주성분분석 요인적재치

〈표Ⅳ-2〉에 따르면 주성분1과 주성분2의 고유값(eigenvalue)이 1보다 크며, 이에 대응하는 주성분이 전체 분산의 92.41%를 설명하고 있다. 따라서 8개의 세이버메트릭스 지수는 2개의 주성분으로 설명이 가능하다고 간주할 수 있다. 각 주성분이 갖는 속성을 확인하기 위하여 〈표Ⅳ-3〉의 요인적재치를 살펴보면, 총 분산의 79.18%를 설명하고 있는 첫 번째 주성분에서는 총 8개의 지수 중 BB/K를 제외한 7개의 지수가 공통성을 보이고 있다. 총 분산의 13.23%를 설명하고 있는 주성분2에서는 BB/K와 OBP만이 양의 값을 갖고있다. 이로부터 주성분1은 타자의 종합적인 능력을, 주성분2는 타자의 선구안 및 출루율과 연관된 능력을 나타내는 성분임을 알 수 있다. 따라서 본 연구에서는 전체 분산의 79.178%를 설명하고 있는 주성분1을 경기력으로 설정하는 것이 가장 타당하다고 판단하여 ‘Performance’ 독립변수로 설정하여 다중회귀분석을 수행하였다.

## (2) 2012 ~ 2016년 KBO리그 타자 다중회귀분석 결과

2012~2016년 KBO리그 타자의 연봉 결정요인을 분석하기 위하여 다중회귀분석을 실시하였다. 해당 시즌 선수의 기록이 다음 시즌 각 선수 연봉에 어떠한 영향을 미치는지를 알아보기 위하여 독립변수는 당 해 선수의 기록으로, 종속변수는 다음 시즌의 연봉으로 분석을 진행하였다. 독립변수의 경우 주성분분석을 통하여 산출한 경기력을 나타내는 'Performance'변수, 올스타 팬 투표 득표수에 따른 'Popularity', 선수의 데뷔시즌을 기준으로 계산한 'Career', 올스타전 출전 횟수/경력으로 나타낸 'Allstar', 'Age'(나이), 'Weight'(몸무게) FA자격구분을 나타내는 'FA', 'Position'(포지션 구분 더미변수), 'Team X'(팀 구분 더미변수), 'Right'(좌타우타 구분)으로 설정하여 선형관계를 살펴보았다. 분석 결과는 <표Ⅳ-4>와 같다.

Coefficients:						
	Estimate	Std. Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-15566.14	0	26669.28	-0.584	0.560115	
Performance	4344.07	0.3601	563.86	7.704	6.51E-13	***
Popularity	-5.05	-0.0043	52.77	-0.096	0.923904	
Career	1786.83	0.2533	646.32	2.765	0.006246	**
Allstar	3663.67	0.2967	784.89	4.668	5.65E-06	***
Age	-1317.24	-0.1810	716.03	-1.84	0.067342	.
Weight	361.68	0.1168	177.09	2.042	0.042464	*
Right	954.42	0.0156	3039.66	0.314	0.753864	
FA	24094.13	0.3868	4260.92	5.655	5.49E-08	***
Position C	10746.11	0.0773	7609.55	1.412	0.159489	
Position 1B	7483.87	0.0784	6153.65	1.216	0.225392	
Position 2B	13237.51	0.1411	6972.21	1.899	0.059092	.
Position 3B	7395.74	0.0814	6271.33	1.179	0.239718	
Position SS	16411.91	0.1658	7793.01	2.106	0.036486	*
Position OF	11348.59	0.1839	5826.36	1.948	0.052874	.
Team1	12049.69	0.1193	7987.13	1.509	0.13301	
Team2	21675.04	0.2271	8171.29	2.653	0.008646	**
Team3	20612.43	0.2082	8559.43	2.408	0.016964	*
Team4	24069.72	0.2522	8053.26	2.989	0.003161	**
Team5	28287.18	0.2963	7746.59	3.652	0.000335	***
Team6	9689.15	0.0979	8340.3	1.162	0.246766	
Team7	19156.77	0.1897	8141.72	2.353	0.019623	*
Team8	33672.95	0.3266	7874.11	4.276	2.97E-05	***
Team9	22110.87	0.2356	8369.96	2.642	0.008919	**
Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1						
Residual standard error: 17010 on 195 degrees of freedom, N: 219						
Multiple R-squared: 0.7204, Adjusted R-squared: 0.6875						
F-statistic: 21.85 on 23 and 195 DF, p-value: < 2.2e-16						

〈표Ⅳ-4〉 2012~2016시즌 KBO 타자 연봉 결정요인 다중회귀분석 결과

〈표Ⅳ-4〉에 따르면 추정된 회귀모형의 설명력을 나타내주는 결정계수(coefficient of determination)는 0.7204, 독립변수의 자유도를

고려하여 산출되는 수정된 결정계수(adjusted coefficient of determination)는 0.6875로 나타났다. 이는 해당 회귀모형의 종속변수의 분산은 투입된 독립변수에 의해 68.75%만큼 설명되어지는 것으로 해석할 수 있다.

타자의 경기력(Performance), 올스타출전횟수/경력(Allstar), FA 세 변수의 경우 0.001 유의수준에서 연봉과 유의한 관계가 있는 것으로 나타났고, 선수의 경력을 나타내는 Career변수의 경우 0.01 유의수준에서 유의한 것으로 나타났다. 이는 주성분분석으로 정의한 Performance변수가 한 단위 증가하면 약 4,334만원의 연봉 상승효과를, 경력대비 올스타전 출전 횟수가 한 단위 증가하면 약 3,663만원의 연봉상승 효과를, FA자격이 갖추어지면 2억 4,094만원의 연봉상승 효과가 있음을 의미한다. 같은 맥락으로 타자의 경력이 1년 늘어날 때마다 1,786만원의 연봉 상승이 이루어지는 것으로 해석된다. 타자 나이의 경우 연봉과 부적인 관계를 보인 반면, 경력은 정적인 관계를 나타냈다. 이러한 결과는 같은 경력의 선수라면 나이가 젊은 선수가 더 좋은 가치 평가를 받는 것으로 해석될 수 있다. 또한 지명타자를 기준 변수로 설정한 타자의 수비 포지션별 연봉 차이는 유격수 포지션만이 유의한 결과를 보였고, 2017년 한국프로야구리그 정규시즌 순위로 구분한 팀 구분에 따른 연봉 차이는 대부분의 팀에서 10위 팀과 연봉 차이를 나타냈다. 타자의 몸무게 변수 또한 0.05 유의수준에서 연봉과 유의한 관계를 나타냈는데, 이는 몸무게와 장타율(SLG)의 상관관계에서 기인한 것으로 판단된다.

반면 올스타전 팬 투표 결과로 계산한 Popularity가 유의한 결과를

보이지 못한 이유는 올스타 팬 투표 후보로 선정된 선수들 자체가 기본적으로 팬들에게 인기가 있는 선수이기 때문인 것으로 판단된다. 이는 올스타 팬 투표 후보군 내에서의 득표수 차이는 연봉에 큰 영향을 미치지 못하는 것을 의미한다.

## 2) 2004 ~ 2006년 KBO리그 타자의 연봉 결정요인

시간의 흐름에 따른 한국프로야구리그 타자의 연봉 결정요인 변화를 파악하기 위하여 2012~2016년 KBO리그 기준 약 10년전인 2004~2006년 KBO리그 타자들을 대상으로 앞선 분석(IV-1-1)과 같은 절차의 분석을 실시하였다.

### (1) 2004 ~ 2006시즌 KBO리그 타자 주성분분석 결과

2004~2006시즌 KBO리그 타자들의 세이버메트릭스 지수로 수행한 주성분분석 결과는 <표IV-5>와 같다. 2004~2006 시즌 타자의 주성분분석 결과에서는 2012~2016시즌의 결과와 달리, 주성분1에서만 1이상의 고유값을 기록하였다. 하지만 해당 주성분분석의 주성분1은 전체 분산의 83.78%를 설명하고 있고, <표IV-6>의 요인적재치를 살펴보면 앞선 <표IV-3>의 결과와 유사하게 주성분1에서 BB/K를 제외한 7개의 지수가 공통적으로 0.8 이상의 값을 나타내고 있다. 따라서 앞선 2012~2016 한국프로야구리그 타자 연봉 결정요인분석과 마찬가지로 주성분1을 'Performance' 변수로 설정하여 분석을 진행하였다.

	주성분1	주성분2	주성분3	주성분4	주성분5	주성분6	주성분7	주성분8
Eigenvalue	6.702	0.891	0.225	0.141	0.026	0.010	0.005	0
% of var.	83.78	11.13	2.81	1.76	0.325	0.128	0.058	0.001
Cumulative % of var.	83.78	94.91	97.72	99.49	99.81	99.94	99.99	100

〈표Ⅳ-5〉 2004~2006시즌 KBO 타자 주성분분석 결과

	wOBA	WAR	RC27	XR27	BB/K	GPA	SLG	OBP
주성분1	0.988	0.805	0.992	0.987	0.554	0.995	0.904	0.933
주성분2	-0.086	-0.168	-0.041	-0.055	0.814	-0.053	-0.344	0.257

〈표Ⅳ-6〉 2004~2006시즌 KBO 타자 주성분분석 요인적재치

## (2) 2004 ~ 2006년 KBO리그 타자 회귀분석 결과

2004~2006년 KBO리그 타자의 연봉 결정요인을 분석하기 위하여 다중 회귀분석을 실시하였다. 해당 시즌 선수의 기록이 다음 시즌 각 선수 연봉에 어떠한 영향을 미치는지를 알아보기 위하여 독립변수는 당 해 선수의 기록으로, 종속변수는 다음 시즌 각 선수의 연봉으로 설정하였다. 독립변수의 경우 주성분분석을 통하여 산출된 주성분1을 경기력을 나타내는 'Performance' 변수로 설정하였고, 나머지 독립변수들 또한 2012~2016 KBO 타자 분석과 동일한 변수로 설정하였다. 분석대상은 2004~2006시즌 동안 규정타석을 충족하면서 모든 독립변수의 값을 갖고 있는 선수로 설정하였다. 그 중 Cook's distance와 스튜던트 잔차(studentized residual) 분석을 통해 이상점(outlier)으로 판단된 3명의 선수를 분석 대상에서 제외하여 총 96명



의 선수를 대상으로 분석을 진행하였다. 분석 결과는 <표Ⅳ-7>와 같다.

Coefficients:						
	Estimate	Std. Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-20493.98	0	19457.67	-1.053	0.295649	
Performance	559.68	0.1053	491.5	1.139	0.258494	
Popularity	8.64	0.0136	54.51	0.158	0.874498	
Career	372	0.1030	596.98	0.623	0.535103	
Allstar	2072.05	0.3638	613.63	3.377	0.001172	**
Age	-460.27	-0.1404	556.02	-0.828	0.410453	
Weight	289.55	0.1705	157.82	1.835	0.070571	.
Right	9560.55	0.3267	2680.49	3.567	0.000638	***
FA	11649.28	0.4158	3028.3	3.847	0.000252	***
Position C	6211.45	0.1478	4371.55	1.421	0.159552	
Position 1B	9837.21	0.2638	3762.26	2.615	0.010815	*
Position 2B	-2521.33	-0.0529	4823.7	-0.523	0.602747	
Position 3B	1340.28	0.0307	4842.66	0.277	0.782732	
Position SS	5704.03	0.1067	5516.13	1.034	0.304474	
Position OF	13736.97	0.4660	3799.08	3.616	0.000544	***
Team1	11388.95	0.3265	4185.2	2.721	0.008103	**
Team2	1350.45	0.0310	5088.1	0.265	0.791429	
Team3	6732.71	0.1602	4788.22	1.406	0.16388	
Team4	7553.11	0.1918	4348.25	1.737	0.08654	.
Team5	5152.91	0.1269	4528.96	1.138	0.258887	
Team6	2993.37	0.0760	4632.11	0.646	0.520135	
Team7	84.45	0.0016	5059.82	0.017	0.986728	
Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1						
Residual standard error: 8276 on 74 degrees of freedom, N: 96						
Multiple R-squared: 0.7267, Adjusted R-squared: 0.6491						
F-statistic: 9.368 on 21 and 74 DF, p-value: 1.843e-13						

<표Ⅳ-7> 2004~2006시즌 KBO 타자 연봉 결정요인 다중회귀분석 결과

<표Ⅳ-7>에 따르면 결정계수는 0.7267, 수정된 결정계수는 0.6491로 나타났다. 이는 해당 회귀모형의 종속변수의 분산은 투입된 독립변수에

의해 64.91%만큼 설명되어지는 것으로 해석된다. 해당 회귀모형은 앞선 2012~2016 시즌 KBO 타자 모형과 투입된 독립변수는 같지만 연봉과 유의한 관계를 보이는 변수에는 차이를 보였다. 본 모형에서는 FA구분변수와 좌타우타 구분 변수, 그리고 포지션 더미변수에서 외야수 변수는 0.001 유의수준에서, ‘올스타 출전 횟수/경력’, 2006년 KBO리그 정규시즌 순위로 더미변수화 한 ‘Team1’ 변수는 0.01 유의수준에서 연봉과 유의한 관계를 나타냈다. 또한 포지션 더미변수의 1루수 변수는 0.05 유의수준에서 연봉과 유의한 관계를 나타냈다.

하지만 ‘Performance’, ‘Popularity’, ‘Career’, ‘Age’ 변수는 연봉과 유의한 관계를 보이지 않았다. 이와 같은 결과에서 주목해야 할 점은 2012~2016시즌 KBO리그 타자 연봉 결정요인 분석과 같은 독립변수가 투입되었음에도 불구하고 결과에서 차이를 보이고 있다는 것이다. 이는 시기에 따라서 KBO 리그 타자의 가치 평가 요소와 그 중요성에는 변화가 있었을 수도 있음을 의미한다. 특히 2012~2016시즌 KBO리그 타자는 경기력에 따른 연봉에 차이를 보였으나, 2004~2006 시즌에는 주성분분석을 통하여 산출한 경기력 변수는 연봉과 유의한 관계를 보이지 않았다. 또한 2012~2016시즌 KBO 타자 회귀모형에서는 선수의 경기력에 따른 연봉차이가 약 4,500만원이지만 2004~2006시즌 KBO 타자 모형에서는 560만원이고 각 변수의 유의성을 나타내주는 t값 또한 2012~2016 KBO 타자 분석 결과와 비교하여 낮은 수치를 나타냈다. 이와 같은 결과는 주성분분석에 투입된 세이버메트릭스 지표들이 2004~2006시즌에는 KBO

타자의 가치를 평가함에 있어 활용되지 않았을 가능성이 크기에 나타난 결과로 보인다. 최경호(2009)의 연구에서는 한국프로야구에서 경기력 평가 시 단순 기록보다 이를 가공한 세이버메트릭스 지수를 사용하는 것이 타당하다고 이야기하고 있다. 해당 연구가 진행된 시점이 2009년임을 감안하면 2004~2006시즌에는 한국프로야구 리그 내에서 세이버메트릭스 지수의 활용이 미비했을 가능성이 높다.

### 3) 2014 ~ 2016시즌 MLB 타자의 연봉 결정요인

본 연구에서는 한국프로야구보다 역사가 깊고, 시장의 규모가 큰 메이저리그에 대한 연봉 결정요인 분석을 통하여 한국프로야구리그 타자와 메이저리그 타자의 연봉 결정요인을 비교 분석하였다. 분석대상은 2014~2016 MLB 시즌 동안 규정타석을 충족한 선수 중 투입된 독립변수의 모든 값을 갖고있는 선수로 표집하였다. 그 중 Cook's distance와 스튜던트 잔차(studentized residual) 분석을 통해 이상점(outlier)으로 판단된 5명의 선수를 분석 대상에서 제외하여 총 415명의 선수를 대상으로 KBO리그 타자 분석과 같은 절차의 분석을 실시하였다.

#### (1) 2014~2016시즌 MLB 타자 주성분분석 결과

2014~2016시즌 MLB 타자의 세이버메트릭스 지수로 실시한 주성분분석 결과는 <표Ⅳ-8>과 같다. 각 주성분의 고유값이 1이상의 값을 갖는 주성분은 첫번째 주성분과 두번째 주성분으로 나타났다. <표 Ⅳ-9>에

따르면 두 주성분 중 8개의 세이버메트릭스 지수 중 BB/K를 제외한 7개가 비슷한 요인적재치를 보이는 주성분1을 ‘Performance’ 변수로 설정하였다. 주성분1은 데이터 전체 분산의 77.51%를 설명하는 것으로 나타났고, 고유값은 6.201, 요인적재치 또한 7개의 지수가 공통성을 보이는 것으로 보아 타자의 전반적인 능력을 나타내기에 적합하다고 판단하였다.

	주성분1	주성분2	주성분3	주성분4	주성분5	주성분6	주성분7	주성분8
Eigenvalue	6.201	1.005	0.495	0.203	0.065	0.027	0.004	0
% of var.	77.51	12.57	6.18	2.54	0.81	0.34	0.05	0.001
Cumulative % of var.	77.51	90.08	96.26	98.70	99.61	99.95	100	100

〈표Ⅳ-8〉 2014~2016시즌 MLB 타자 주성분분석 결과

	wOBA	WAR	RC27	XR27	BB/K	GPA	SLG	OBP
주성분1	0.989	0.745	0.913	0.985	0.539	0.992	0.893	0.891
주성분2	-0.019	-0.054	-0.325	-0.031	0.803	-0.024	-0.366	0.341

〈표Ⅳ-9〉 2014~2016시즌 MLB 타자 주성분분석 요인적재치

## (2) 2014~2016시즌 MLB 타자 다중회귀분석 결과

2014~2016년 MLB 타자의 연봉 결정요인을 분석하기 위하여 다중회귀 분석을 실시하였다. 독립변수로 투입된 선수의 기록이 다음 시즌 연봉에 어떠한 영향을 미치는지에 대한 분석을 진행하기 위하여, 각 선수의 기록을 기준으로 다음 시즌 연봉을 종속변수로 설정하였다. 독립변수는 주성분분석

을 통하여 산출된 주성분1을 경기력 변수(Performance)로 설정하였고, 선수 경력(Career), 나이(Age), 몸무게(Weight), FA유무(FA), 좌타우타구분(Right), 지명타자를 기준으로 설정한 포지션 더미변수(positionX), 아메리칸리그, 내셔널리그 구분(AL/NL)으로 설정하였다.

KBO 타자의 연봉 결정요인에 대한 분석에서는 올스타 팬 투표 득표에 따른 인기도변수(Popularity)와 (올스타 출전 횟수/경력)으로 계산한 올스타 변수(Allstar)를 독립변수로 설정하였으나, 2014~2016시즌 MLB 타자 다중 회귀분석에서는 자료 수집의 한계로 인하여 올스타 출전 횟수 관련 변수를 회귀모형에 설정하지 못 하였다. 또한 한국프로야구 타자의 연봉 결정요인 분석에서는 독립변수로 설정하였던 팀 구분의 경우 MLB 타자 분석에서도 진행하였으나 유의하지 않게 나타났고, 전체 팀의 수가 30개이기에 R-Square가 지나치게 증가하는 경향을 보여 해당 분석 결과에는 제시하지 않고 아메리칸 리그와 내셔널리그를 구분한 독립변수만 추가하여 진행하였다. 구체적인 분석 결과는 <표 IV-10>와 같다.

Coefficients:						
	Estimate	Std. Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-8678874	0	4513348	-1.923	0.055196	.
Performance	472022	0.1721	91190	5.176	3.59E-07	***
Career	1323257	0.7440	124664	10.615	< 2e-16	***
Age	-222054	-0.1203	125970	-1.763	0.078704	.
Weight	105808	0.1356	28779	3.677	0.000269	***
FA	2066502	0.1439	606209	3.409	0.000718	***
Right	624900	0.0456	446212	1.4	0.162151	
Position C	2639603	0.0952	1310962	2.013	0.044731	*
Position 1B	5142609	0.2531	1145462	4.49	9.34E-06	***
Position 2B	3010941	0.1421	1239276	2.43	0.015554	*
Position 3B	2950626	0.1518	1186386	2.487	0.013286	*
Position SS	3439765	0.1732	1276137	2.695	0.007325	**
Position OF	4047651	0.2827	1114548	3.632	0.000318	***
AL/NL	169688	0.0123	432398	0.392	0.694946	
Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1						
Residual standard error: 4227000 on 401 degrees of freedom, N: 415						
Multiple R-squared: 0.6305, Adjusted R-squared: 0.6186						
F-statistic: 52.64 on 13 and 401 DF, p-value: < 2.2e-16						

〈표Ⅳ-10〉 2014~2016시즌 MLB 타자 연봉 결정요인 다중회귀분석 결과

〈표Ⅳ-10〉에 따르면 결정계수는 0.6305, 수정된 결정계수는 0.6186으로 나타났다. 이를 통해 2014~2016시즌 메이저리그 타자 연봉은 해당 모형에 투입된 독립변수에 의해 61.86%만큼 설명되는 것으로 해석할 수 있다. 각각의 독립변수의 유의성을 살펴보면, 경기력(Performance), 경력(Career), FA유무(FA), 몸무게(Weight) 지명타자가 기준인 포지션 터미변수에서는 1루수(Position3)와 외야수(Positon OF) 변수가 0.001 유의수준에서 연봉과 유의한 관계를 보이는것으로 나타났다. 유격수(Position SS)는 0.01 유의수준에서 연봉과 유의한 관계를 나타냈고, 나머지 포수(Position C),

2루수(Position 2B), 3루수(Position 3B) 포지션 또한 0.05 유의수준에서 연봉에 유의한 관계를 나타냈다.

분석 결과를 요약하면 다음과 같다. 메이저리그에서는 경기력, 경력, FA유무, 몸무게, 포지션 구분 등 다양한 변수가 연봉에 영향을 미치는 것으로 나타났다. 특히 한국프로야구 타자의 경우 포지션 더미변수에서 소수의 포지션만 유의성을 나타냈으나 해당 모형에서는 모든 포지션에서 유의성을 나타냈다. 이와 같은 결과는 각 리그에서의 수비 능력에 대한 가치 평가가 다를 수 있음을 보여주는 결과로 사료된다. 또한 표준화 회귀계수를 살펴보면 선수 경력이 가장 높은 값을 나타내고 있는데, 이는 세계 최고 수준이라고 할 수 있는 메이저리그에서 계속 선수 생활을 했다면, 높은 가치 평가가 이루어지고 있음을 나타내주는 결과로 사료된다.

## 2. 투수의 연봉 결정요인 분석

### 1) 2012~2016시즌 KBO리그 투수의 연봉 결정요인

2012~2016시즌 KBO리그 투수의 연봉 결정요인을 탐색하기 위하여 주성분분석과 다중회귀분석을 실시하였다. 2012~2016시즌 KBO 투수의 연봉 결정요인 다중회귀모형은 주성분분석을 통해 산출된 투수의 투구능력 변수(performance1)와 투수의 승리 관련 기록 변수(performance2), 그리고 이외의 다양한 요인들을 독립변수로 설정하여 연봉과의 선형관계를 살펴보았다. 분석대상은 2012~2016시즌 동안 규정이닝을 충족한 투수 중 투입된 독립변수의 모든 값을 갖고 있는 선수 중 Cook's distance와 스튜던트 잔차(studentized residual) 분석을 통해 이상점(outlier)으로 판단된 3명의 선수를 분석 대상에서 제외하여 총 86명의 선수를 대상으로 분석을 진행하였다. 세부적인 분석 결과는 다음과 같다.

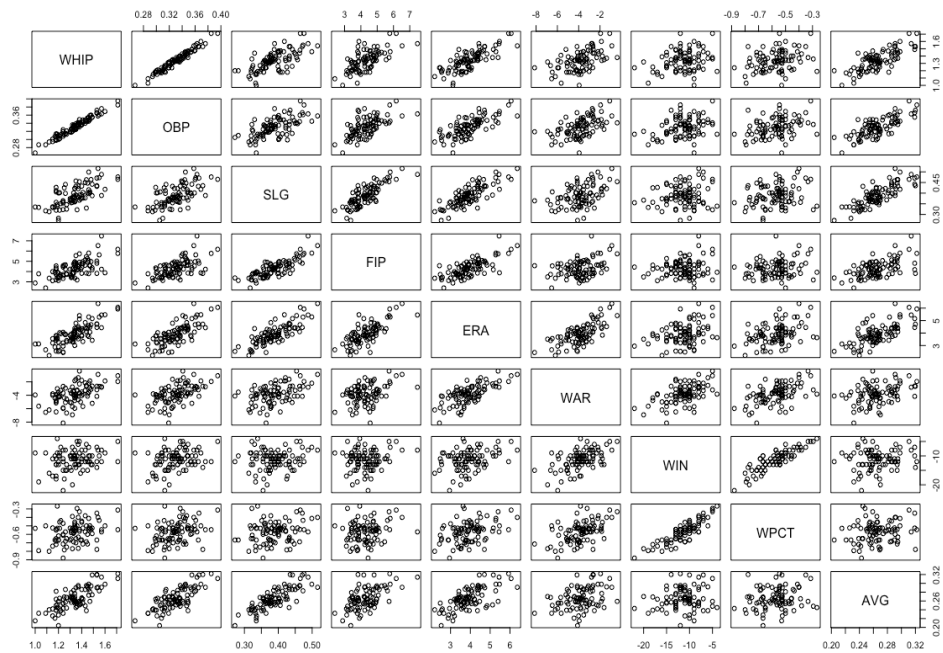
#### (1) 2012~2016시즌 KBO리그 투수 주성분분석 결과

주성분분석을 위한 데이터가 주성분분석의 기본 가정을 만족하는지 확인하기 위하여 Bartlett의 구형성검정을 수행하였다. 분석 결과 검정통계량 값이 2746.8( $p < 0.001$ )로 주성분분석의 가정을 만족하는 것으로 나타났다. 이에 따라서 투수의 경기력을 나타내주는 9가지 세이버메트릭스 지수의 차원축소를 위하여 주성분분석을 실시하였다. 투수 기록의 경우 선수의 경기력과 음의 상관관계를 갖는 기록(ERA, FIP, AVG,



WHIP 등)과 양의 상관관계를 갖는 기록(WAR, Win, WPCT 등)으로 나누어진다. 이에 대하여 요인적재치의 동일한 방향 설정을 위해 주성분분석에 투입된 9가지 세이버메트릭스 지수는 낮은 값을 갖는 경우 우수한 경기력을 나타내는 것으로 나타나도록, 양의 상관관계를 갖는 기록에는 -1을 곱하였다. 이후 산출된 주성분의 각각의 값에 다시 -1을 곱하여 높은 수치가 우수한 경기력을 나타내도록 코딩하였다.

〈그림 IV-2〉는 해당 주성분분석에 투입된 세이버메트릭스 지수들의 상관관계를 산점도로 나타낸 것이다. 정확한 상관계수는 〈표 IV-11〉와 같다. 〈표 IV-11〉를 살펴보면, 상관계수의 최대값은 0.9768, 최저값은 0.1262로 9가지의 세이버메트릭스 지수 중 하나의 지수를 경기력이라고 설정하기 어려움을 알 수 있다. 이에 본 연구에서는 주성분분석을 통하여 투수의 경기력을 투구능력(performance1)과 승리관련 기록(performance2)으로 나누어 변수로 설정하였다. 주성분분석 결과는 〈표 IV-12〉와 같다.



〈그림 IV-2〉 2012~2016년 KBO리그 투수 세이버메트릭스 지표들의 상관관계 산점도

	WHIP	OBP	SLG	FIP	ERA	WAR	WIN	WPCT	AVG
WHIP	1.0000								
OBP	0.9768	1.0000							
SLG	0.6813	0.6390	1.0000						
FIP	0.6597	0.6331	0.8129	1.0000					
ERA	0.7760	0.7333	0.8094	0.7307	1.0000				
WAR	0.5345	0.5298	0.4550	0.3761	0.7013	1.0000			
WIN	0.2579	0.2974	0.1285	0.1364	0.3370	0.5487	1.0000		
WPCT	0.2960	0.3324	0.2290	0.1872	0.4089	0.5205	0.8697	1.0000	
AVG	0.7696	0.7614	0.8143	0.5654	0.6853	0.3704	0.1262	0.1784	1.0000

〈표 IV-11〉 2012~2016시즌 KBO 투수 세이버메트릭스 지표 상관계수

〈표 IV-12〉의 결과에 따르면 주성분1과 주성분2의 고유값(eigenvalue)이 1보다 크며, 이에 대응하는 주성분이 전체 분산의 79.862%를 설명하고 있다. 〈표 IV-13〉을 살펴보면, 총 분산의 59.624%를 설명하고 있는 첫 번째

주성분에서는 총 9개의 지수 중 승리수(WIN)와, 승리확률(WPCT)을 제외한 7개의 지수가 공통성을 보이고 있다. 총 분산의 20.24%를 설명하고 있는 주성분2에서는 투수의 승리와 관련된 기록인 대체 선수 대비 승리기여도(WAR), 승리수(WIN)와 승리확률(WPCT)만이 양의 값을 갖고있다. 이는 주성분1은 투수의 투구 능력을 나타내주고 있고, 주성분2는 투수의 승리 관련 기록을 나타내주는 성분인 것으로 간주할 수 있다. 본 연구에서는 전체 분산의 59.62%를 설명하고 있는 주성분1을 2012~2016시즌 KBO투수 회귀분석모형의 ‘Performance1’ 변수로, 주성분2를 ‘Performance2’ 변수로 설정하여 다중회귀분석을 수행하였다.

	주성분1	주성분2	주성분3	주성분4	주성분5	주성분6	주성분7	주성분8	주성분9
Eigenvalue	5.366	1.821	0.600	0.504	0.375	0.157	0.107	0.052	0.018
% of var.	59.62	20.24	6.67	5.60	4.16	1.74	1.19	0.58	0.19
Cumulative % of var.	59.62	79.86	86.53	92.13	96.29	98.03	99.22	99.81	100

〈표 IV-12〉 2012~2016시즌 KBO 투수 주성분분석 결과

	WHIP	OBP	SLG	FIP	ERA	WAR	WIN	WPCT	AVG
주성분1	0.901	0.888	0.849	0.777	0.918	0.705	0.445	0.498	0.810
주성분2	-0.169	-0.122	-0.310	-0.291	-0.032	0.394	0.842	0.790	-0.328

〈표 IV-13〉 2012~2016시즌 KBO 투수 주성분분석 요인적재치

## (2) 2012~2016시즌 KBO리그 투수 다중회귀분석 결과

2012~2016시즌 KBO 투수의 연봉 결정요인을 분석하기 위하여 다중회귀분석을 실시하였다. 당 해 선수의 기록이 다음 시즌 각 선수 연봉에 어떠한 영향을 미치는지를 알아보기 위하여 독립변수는 당 해 선수의 기록으로, 종속변수는 다음 시즌 각 선수 연봉으로 설정하였다. 독립변수는 주성분분석을 통하여 산출된 주성분1을 투수의 투구 능력에 대한 변수(Performance1)로 설정하였고, 주성분2를 투수의 승리 관련 기록(Performance2)로 설정하였다. 또한 나이(Age), 올스타전 출전 횟수(Allstar), FA자격구분(FA), 키(Height), 좌투우투구분(Right), 용병구분(foreign), 팀 구분(Team1~9)을 독립변수로 설정하였다. 팀 구분의 경우 2017년 KBO리그 정규시즌 순위에 따라서 Team1~10으로 더미변수화하였고, 10개 팀 중 2017년 KBO리그 정규시즌 10위 팀을 더미변수의 기준변수 설정하여 다중회귀분석을 실시하였다. 구체적인 결과는 <표 IV-14>와 같다.

Coefficients:						
	Estimate	Std. Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-63416.54	0	109391.51	-0.58	0.564047	
Performance1	64.59	0.0052	1094.74	0.059	0.953129	
Performance2	8025.2	0.3653	1974.76	4.064	0.000129	***
Career	3539.94	0.4223	1311.05	2.7	0.008768	**
Age	-1937.75	-0.2230	885.02	-2.189	0.032049	*
Allstar	8214.02	0.1915	4016.39	2.045	0.044773	*
Height	613.44	0.1070	568.08	1.080	0.284083	
FA	29282.63	0.5033	9303.75	3.147	0.002459	**
Right	396.34	0.0068	5573.02	0.071	0.943516	
Foreign	11335.81	0.1941	13185.55	0.86	0.393012	
Team1	15042.94	0.1517	16999.33	0.885	0.379368	
Team2	28887.19	0.3474	16260.26	1.777	0.080182	.
Team3	12538.73	0.1508	16030.12	0.782	0.436854	
Team4	12864.50	0.1221	17466.04	0.737	0.463971	
Team5	12476.39	0.1326	16590.51	0.752	0.454673	
Team6	12346.21	0.1431	16473.09	0.749	0.456193	
Team7	-1872.98	-0.0199	16087.10	-0.116	0.907662	
Team8	-9414.51	-0.0949	17342.90	-0.543	0.589038	
Team9	18871.93	0.1902	17493.49	1.079	0.284547	
Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1						
Residual standard error: 19850 on 67 degrees of freedom, N: 86						
Multiple R-squared: 0.6301, Adjusted R-squared: 0.5308						
F-statistic: 6.342 on 18 and 67 DF, p-value: 1.031e-08						

〈표Ⅳ-14〉 2012~2016시즌 KBO 투수 연봉 결정요인 다중회귀분석 결과

〈표Ⅳ-14〉에 따르면 결정계수는 0.6301, 수정된 결정계수는 0.5308로 나타났다. 이는 2014~2016시즌 KBO리그 투수 연봉의 분산은 해당 모형의 독립변수에 의해 53.08%만큼 설명되는 것으로 해석할 수 있다. 각각의 독립변수의 유의성을 살펴보면, 투수의 승리 관련 기록을 나타내는 ‘Performance2’ 변수의 유의확률은 0.001보다 작게 나타나 연봉과 매우

유의한 관계를 보였다. 선수의 경력을 나타내는 ‘Career’변수와 FA구분 변수는 0.01 유의수준에서 연봉과 유의한 관계를 보였다. 또한 선수의 나이(Age)와 올스타전 출전 횟수(Allstar)는 0.05 유의수준에서 연봉과 유의한 관계가 있는 것으로 나타났다. 하지만 투수의 투구 능력을 나타내는 ‘Performance1’ 변수는 통계적으로 연봉과 유의한 영향을 미치지 않는 것으로 나타났고, 2017년 KBO리그 정규시즌 순위로 설정한 팀 구분 더미변수(Team X) 또한 유의하지 않았다.

해당 회귀모형에서 주목해야 할 부분은 선수의 가치 평가시 개인의 투구 능력을 나타내주는 ‘Performance1’ 변수보다 팀의 득점 생산 능력, 수비 능력에 따라 달라질 수 있는 투수의 승리 관련 기록인 ‘Performance2’에서 연봉에 더욱 유의한 영향을 미치는 것으로 나타났다는 것이다. 회귀계수 또한 ‘Performance1’ 변수는 한 단위 증가하면 연봉이 1,502만원, ‘Performance2’변수의 한 단위 증가는 8,584만원의 연봉이 상승하는 것으로 나타났다. 이는 선수 개인의 가치를 평가하는 데에 있어 개인의 능력도 중요하지만, 팀 구분에 따른 승리에 대한 운이 낮다면 타당한 가치 평가가 이루어지지 않을 수 있음을 시사한다.

## 2) 2004~2006시즌 KBO리그 투수의 연봉 결정요인

시간의 흐름에 따른 한국프로야구리그 투수의 연봉 결정요인 변화를 파악하기 위하여 2012~2016년 KBO리그 기준으로 약 10년전인 2004~2006년시즌 KBO리그 투수들을 대상으로 주성분분석과 다중회귀

분석을 실시하였다. 분석대상은 2004~2006시즌 동안 규정이닝을 충족한 선수 중 투입된 독립변수의 모든 값을 갖고 있는 선수로 설정하였다. 그 중 Cook's distance와 스튜던트 잔차(studentized residual)를 통해 이상점(outlier)으로 판단된 2명의 선수를 분석 대상에서 제외하여 52명의 선수로 설정하였다. 세부적인 분석 결과는 다음과 같다.

#### (1) 2004~2006시즌 KBO리그 투수 주성분분석 결과

투수의 경기력을 나타내는 9가지 세이버메트릭스 지수의 차원축소를 위하여 주성분분석을 실시하였다. 투수 기록의 경우 선수의 경기력과 음의 상관관계를 갖는 기록(ERA, FIP, AVG, WHIP 등)과 양의 상관관계를 갖는 기록(WAR, Win, WPCT 등)으로 나뉘어진다. 이에 대하여 요인적재치의 동일한 방향 설정을 위해 주성분분석에 투입된 9가지 세이버메트릭스 지수는 낮은 값을 갖는 경우 우수한 경기력을 나타내는 것으로 나타나도록, 양의 상관관계를 갖는 기록에는 -1을 곱하였다. 이후 산출된 주성분 벡터의 각각의 값에 다시 -1을 곱하여 높은 수치가 우수한 경기력을 나타내도록 코딩하였다.

분석에 앞서 해당 데이터가 주성분분석의 기본 가정을 만족하는지 확인하기 위하여 Bartlett의 구형성검정을 수행하였다. 분석 결과 검정통계량 값이 1726.7 ( $p < 0.001$ )로 주성분분석의 가정을 만족하는 것으로 나타나 주성분분석을 시행하였으며, 그 결과는 <표 IV-15>와 같다.

	주성분1	주성분2	주성분3	주성분4	주성분5	주성분6	주성분7	주성분8	주성분9
Eigenvalue	5.978	1.399	0.646	0.392	0.264	0.155	0.102	0.046	0.018
% of var.	66.420	15.55	7.18	4.36	2.93	1.73	1.13	0.51	0.20
Cumulative % of var.	66.42	81.97	89.15	93.50	96.43	98.16	99.29	99.80	100

〈표Ⅳ-15〉 2004~2006시즌 KBO 투수 주성분분석 결과

	WHIP	OBP	SLG	FIP	ERA	WAR	WIN	WPCT	AVG
주성분 1	0.869	0.863	0.865	0.839	0.928	0.843	0.640	0.603	0.825
주성분 2	-0.327	-0.312	-0.105	-0.210	-0.026	0.235	0.697	0.720	-0.282

〈표Ⅳ-16〉 2004~2006시즌 KBO 투수 주성분분석 요인적재치

〈표Ⅳ-15〉에 따르면 주성분1과 주성분2의 고유값 (eigenvalue)이 1보다 크며, 이에 대응하는 주성분이 전체 분산의 81.97%를 설명하고 있다. 〈표Ⅳ-16〉의 요인적재치를 살펴보면, 총 분산의 66.42%를 설명하고 있는 주성분1에서는 총 9개의 지표 중 승리수(WIN)와, 승리확률(WPCT)을 제외한 7개의 지표가 공통성을 보이고 있다. 총 분산의 15.55%를 설명하고 있는 주성분2에서는 투수의 승리와 관련된 기록인 대체 선수 대비 승리기여도(WAR), 승리수(WIN)와 승리확률(WPCT)만이 양의 값을 갖고있다. 이는 앞선 2012~2016 KBO투수를 대상으로 실시한 주성분분석 결과와 마찬가지로 주성분1은 투수의 투구 능력을 나타내고 있으며 주성분2는 투수의 승리 관련 기록을 나타내는 성분임을 알 수 있다. 본 연구에서는 전체 분산의 66.42%를 설명하고 있는 주성분1을 2004~2006시즌 KBO투수 회귀분석모형의 ‘Performance1’ 변수로, 주성분2를 ‘Performance2’ 변수로 설정하여 다중회귀분석을 수행하였다.



## (2) 2004~2006시즌 KBO리그 투수 다중회귀분석 결과

시간의 흐름에 따른 KBO 투수 연봉 결정요인의 변화를 확인하기 위하여 다중회귀분석을 실시하였다. 분석 대상은 2004~2006시즌 한국프로야구 투수 중 각 시즌 규정이닝을 충족하면서 투입된 독립변수의 모든 값을 갖고있는 선수로 설정하였다. 종속변수는 독립변수 기록시즌을 기준으로 다음 시즌 각 선수의 연봉으로 설정하였다. 독립변수는 주성분분석을 통하여 산출된 주성분1을 투수의 투구 능력 변수(performance1)로 설정하였고, 주성분2를 투수의 승리 관련 기록 변수(performance2)로 설정하였다. 또한 선수경력(Career), 나이(Age), 올스타출전횟수/경력(Allstar), FA유무(FA), 키(Height), 좌투우투구분(Right), 용병구분(foreign), 팀 구분(Team1~9)변수도 함께 독립변수로 설정하였다. 세부적인 분석 결과는 <표 IV-17>와 같다.

Coefficients:						
	Estimate	Std. Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	87941.5	0	37495	2.345	0.0248	*
Performance1	1576.4	0.3688	309	5.102	1.18E-05	***
Performance2	504.3	0.0587	613.2	0.822	0.41641	
Career	1112.7	0.5210	396.6	2.805	0.00815	**
Age	-403.1	-0.2046	321.1	-1.256	0.2176	
Allstar	3379.3	0.3086	1230.5	2.746	0.00945	**
Height	-411.6	-0.1710	196.8	-2.092	0.04381	*
FA	5848.4	0.2732	2759.8	2.119	0.04125	*
Right	6294.9	0.2310	2045.2	3.078	0.00404	**
Foreign	10775.1	0.3955	4570.4	2.358	0.02411	*
Team1	6860.4	0.2401	2943.2	2.331	0.02564	*
Team2	-2314	-0.0849	2977.8	-0.777	0.44235	
Team3	2866.9	0.0949	2870.8	0.999	0.32482	
Team4	-1926.2	-0.0498	3382.7	-0.569	0.5727	
Team5	-4403.5	-0.1365	3514.4	-1.253	0.21851	
Team6	-2572.6	-0.0852	3393.4	-0.758	0.45346	
Team7	2501	0.0078	3313.7	0.755	0.45544	
Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1						
Residual standard error: 4653 on 35 degrees of freedom, N: 52						
Multiple R-squared: 0.8628, Adjusted R-squared: 0.8001						
F-statistic: 13.75 on 16 and 35 DF, p-value: 1.241e-10						

〈표Ⅳ-17〉 2004~2006시즌 KBO 투수 연봉 결정요인 다중회귀분석 결과

〈표Ⅳ-17〉에 따르면 결정계수는 0.8628, 수정된 결정계수는 0.8001로 나타났다. 이는 2004~2006시즌 KBO 투수 연봉의 분산은 해당 모형의 독립변수에 의해 80.1%만큼 설명되어지는 것을 의미한다. 2012~2016시즌 KBO리그 투수 연봉 결정요인 모형과 같은 독립변수를 투입하였음에도 해당 모형의 결정계수가 높게 나타난 것으로 보아 2004~2006시즌

KBO리그 투수의 연봉 결정요인은 2012~2016시즌 KBO리그 투수의 연봉 결정요인에 비하여 더 적은 수의 요인에 의해 결정되었던 것으로 보인다.

각각의 독립변수에서는 투수의 투구 능력을 나타내주는 'Performance1' 변수의 유의확률은 0.001보다 작게 나타났고 'Career', 'Right', 'Allstar' 변수는 0.01 유의수준에서 연봉과 유의한 관계를 보였다. 또한 'Height', 'FA', 'Foreign', 'Team1' 변수는 유의수준 0.05에서 유의한 것으로 나타났다.

해당 모형을 2012~2016시즌 KBO 투수의 연봉 결정요인 모형과 비교하여 살펴보면, 2012~2016시즌 KBO 투수의 연봉 결정요인은 승리 관련 기록 변수(Performance2)가 가장 낮은 유의확률을 기록하였으나, 2004~2006시즌 KBO투수의 경우 투수의 투구 능력을 나타내는 변수가 연봉과 가장 유의한 관계를 보이는 것으로 나타났다. 또한 선수 경력과 올스타 출전 횟수, FA유무는 공통적으로 연봉에 유의한 영향을 미치는 것으로 나타났다.

### 3) 2014~2016시즌 MLB 투수의 연봉 결정요인

한국프로야구리그 투수의 연봉 결정요인과 메이저리그 투수의 연봉 결정요인의 비교 분석을 위하여 2014~2016시즌 MLB 투수를 대상으로 KBO리그 투수 분석과 같은 절차의 분석을 실시하였다. 분석대상은 2014~2016 시즌 동안 규정이닝을 충족한 선수 중 투입된 독립변수의 모든 값을 갖고있는 선수로 설정하였다. 그 중 Cook's distance와 스튜던트 잔차(Studentized residual)를 통해 이상점(outlier)으로 판단된 1명의 선수를

분석 대상에서 제외하여 총 220명의 선수를 대상으로 분석을 진행하였다.

#### (1) 2014~2016시즌 메이저리그 투수 주성분분석 결과

투수의 경기력을 나타내는 9가지 세이버메트릭스 지수의 차원축소를 위하여 주성분분석을 실시하였다. 투수의 기록의 경우 선수의 경기력과 음의 상관관계를 갖는 기록(ERA, FIP, AVG, WHIP 등)과 양의 상관관계를 갖는 기록(WAR, Win, WPCT 등)으로 나뉘어진다. 이에 대하여 요인적재치의 동일한 방향 설정을 위해 주성분분석에 투입된 9가지 세이버메트릭스 지수는 낮은 값을 갖는 경우 우수한 경기력을 나타내는 것으로 나타나도록, 양의 상관관계를 갖는 기록에는 -1을 곱하였다. 이후 산출된 주성분 벡터의 각각의 값에 다시 -1을 곱하여 높은 수치가 우수한 경기력을 갖도록 코딩하였다.

분석에 앞서 해당 데이터가 주성분분석의 기본 가정을 만족하는지 확인하기 위하여 Bartlett의 구형성검정을 수행하였다. 분석결과 검정통계량 값은 7241.6, 유의확률은 0.001보다 작은것으로 나타나 주성분분석의 가정을 만족하는 것으로 나타났다. 이에 따라 주성분분석을 시행하였으며 결과는 <표 IV-18>과 같다.

	주성분1	주성분2	주성분3	주성분4	주성분5	주성분6	주성분7	주성분8	주성분9
Eigenvalue	6.271	1.010	0.628	0.437	0.294	0.148	0.132	0.067	0.013
% of var.	69.68	11.22	6.97	4.86	3.27	1.65	1.46	0.74	0.15
Cumulative % of var.	69.68	80.90	87.87	92.73	96.0	97.65	99.11	99.85	100

〈표 IV-18〉 2014~2016시즌 MLB 투수 주성분분석 결과

	WHIP	OBP	SLG	FIP	ERA	WAR	WIN	WPCT	AVG
주성분 1	0.918	0.908	0.839	0.737	0.911	0.886	0.708	0.733	0.839
주성분 2	-0.108	-0.111	-0.253	-0.248	-0.151	-0.041	0.647	0.622	-0.173

〈표 IV-19〉 2014~2016시즌 MLB 투수 주성분분석 요인적재치

위 표를 구체적으로 살펴보면 주성분1과 주성분2의 고유값(eigenvalue)이 1보다 크며, 이에 대응하는 두 개의 주성분이 전체 분산의 80.90%를 설명하고 있다. 각 주성분의 요인적재치를 살펴보면, 총 분산의 69.68%를 설명하고 있는 주성분1에서는 총 9개의 지수 중 승리수(WIN)와, 승리확률(WPCT), 투수가 통제가능한 영역만으로 투수를 평가하는 지수인 'FIP' 세 지수가 0.75이하의 값을 가지고 있고, 나머지 6개의 지수는 0.8이상의 값을 나타내고 있다. 한국프로야구리그 투수의 주성분분석 결과에서는 승리수(Win)와 승리확률(WPCT)만이 낮은 요인적재치를 나타냈으나 메이저리그의 경우 주성분1에서 9개의 변수 모두 0.7이상의 요인적재치를 나타냈다. 이와 같은 차이는 투수의 우수한 경기력이 승리로 이어질 확률이 한국프로야구리그보다 메이저리그에서 높다는 리그 특성에서 기인하는 결과로 사료된다. 총 분산의 11.223%를 설명하고 있는

주성분2에서는 승리수(WIN)와 승리확률(WPCT)만이 양의 값을 기록하였다. 한국프로야구 리그 투수 주성분분석에서는 대체 선수 대비 승리 기여도 (WAR) 또한 주성분2에서 양의 값을 나타냈으나, 해당 분석에서는 -0.041로 0에 가까운 값을 나타냈다. 이와 같이 요인적재치의 검토에 따라서 주성분1은 투수의 투구 능력을 나타내고 있으며 주성분2는 투수의 승리 관련 기록을 나타내는 성분임을 알 수 있다.

본 연구에서는 전체 분산의 69.68%를 설명하고 있는 주성분1을 2014~2016시즌 MLB 투수 회귀분석모형의 'Performance1' 변수로, 주성분2를 'Performance2' 변수로 설정하여 다중회귀분석을 수행하였다.

## (2) 2014~2016시즌 메이저리그 투수 다중회귀분석 결과

한국프로야구리그와 메이저리그 투수의 연봉 결정요인을 비교 분석하기 위하여 다중회귀분석을 실시하였다. 분석 대상은 2014~2016시즌 동안 규정이닝을 충족한 투수 중 투입된 독립변수의 모든 값을 갖고있는 선수로 설정하였다. 그 중 Cook's distance와 스튜던트 잔차(studentized residual)를 통해 이상점(outlier)으로 판단된 1명의 선수를 분석 대상에서 제외하여 총 220명의 선수를 대상으로 분석을 진행하였다. 독립변수는 주성분분석을 통하여 산출된 주성분1을 투수의 투구 능력 변수(Performance1)로 설정하였고, 주성분2를 투수의 승리 관련 기록 변수(Performance2)로 설정하였다. 또한 나이(Age), 선수 경력(Career), 올스타출전횟수/경력 (Allstar), FA자격구분(FA), 키(Height), 좌투우투구분(Right), 아메리칸리그,

내셔널리그의 구분(AL/NL) 변수 또한 독립변수로 설정하여 다중회귀 분석을 진행하였다. 구체적인 분석 결과는 <표Ⅳ-20>과 같다.

Coefficients:						
	Estimate	Std. Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	12153877	0	11957156	1.016	0.310584	
Performance1	566926	0.1847	155935	3.636	0.000349	***
Performance2	542160	0.0687	353424	1.534	0.12653	
Career	1280117	0.6113	206318	6.205	2.88E-09	***
Age	-320247	-0.1622	191168	-1.675	0.095381	.
Allstar	15925767	0.3453	2569363	6.198	2.98E-09	***
Height	-22990	-0.017	60823	-0.378	0.705827	
FA	1884755	0.1182	948724	1.987	0.048265	*
Right	-1709751	-0.0984	810036	-2.111	0.03598	*
AL/NL	51773	0.0033	694584	0.075	0.940653	
Signif. codes: '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1						
Residual standard error: 5068000 on 210 degrees of freedom, N: 220						
Multiple R-squared: 0.5948, Adjusted R-squared: 0.5774						
F-statistic: 34.25 on 9 and 210 DF, p-value: < 2.2e-16						

<표Ⅳ-20> 2014~2016시즌 MLB 투수 연봉 결정요인 다중회귀분석 결과

<표 Ⅳ-20>에 따르면 결정계수는 0.5948, 수정된 결정계수는 0.5774로 나타났다. 따라서 2014~2016시즌 MLB 투수 연봉의 분산은 해당 모형의 독립변수에 의해 57.74%만큼 설명되어지는 것으로 해석 할 수 있다.

각 독립변수의 연봉에 대한 유의성을 살펴보면, 투수의 투구 능력을 나타내주는 'Performance1', 선수의 경력(Career), 올스타 출전 횟수/경력(Allstar)변수의 유의확률은 0.001보다 작게 나타났다. 또한 FA 자격 구분 변수(FA)와 좌투우투 구분 변수(Right)는 0.05 유의수준에서 연봉에 유의한

영향을 미치는 것으로 나타났다.

해당 모형을 2012~2016 KBO 투수의 연봉 결정요인 모형과 비교하여 살펴보면, 2012~2016시즌 KBO투수는 FA자격 구분과 승리 관련 기록에서 0.001보다 작은 유의확률을 기록하면서 표준화회귀계수 또한 다른 독립변수들과 비교하여 상대적으로 큰 값을 나타냈다. 반면 2014~2016시즌 MLB투수 모형에서는 선수 경력 변수가 가장 큰 표준화 회귀계수를 보였고, 유의확률 또한 가장 낮게 나타났다. 또한 선수의 승리관련 기록이 아닌 선수의 투구능력을 나타내는 ‘Performance1’ 변수가 연봉과 유의한 영향을 미친다는 점에서 2012~2016시즌 KBO리그 투수 연봉 결정요인과 차이를 보였다. 이는 2014~2016시즌 MLB 투수의 연봉은 선수의 승리에 대한 기록보다는 선수 개인의 투구 능력에 대한 기록을 중요시하고, 세계 최고 수준의 경기력을 보이는 MLB에서의 오랜 선수 경력과 선수의 인기로 대변될 수 있는 올스타전 출전 기록은 선수의 가치를 평가함에 있어 중요한 요인으로 간주할 수 있을 것이다.



## V. 논의

본 연구의 목적은 한국프로야구리그와 메이저리그 선수들의 연봉 결정요인을 탐색하고, 분석 결과를 토대로 한국프로야구리그의 연봉 시장을 진단하며, 한국프로야구리그 특성에 맞는 연봉 책정 지표 개발을 위한 기초자료를 제시하는 것이다. 이러한 연구목적에 따라 2012~2016시즌 KBO리그, 2004~2006시즌 KBO리그, 2014~2016시즌 MLB리그 선수의 연봉 결정요인 분석을 총 여섯개의 다중회귀모형을 설정하여 분석을 진행하였다. 분석 결과를 바탕으로한 논의 내용은 다음과 같다.

### 1. 타자

본 연구에서는 프로야구 타자 연봉 결정요인을 분석하기 위하여 2012~2016 시즌 KBO 타자 회귀모형, 2004~2006 시즌 KBO 타자 회귀모형, 2014~2016 시즌 MLB 타자 회귀모형을 설정하여 분석을 진행하였다. 세 모형 모두 공통적으로 FA 자격 구분 변수에서 연봉과 통계적으로 유의한 선형관계를 나타냈다. 반면 선수의 경기력 변수는 2012~2016 시즌 KBO 타자 모형과 2014~2016 시즌 MLB 타자 모형에서는 유의한 결과를 보였으나, 2004~2006 시즌 KBO 타자의 연봉 결정요인 분석에서는 유의성을 보이지 않았다. 이와 같은 결과는

2004~2006 시즌 KBO 리그에서는 세이베메트릭스 지수가 많이 쓰이지 않았고, 단순 기록을 주로 활용하였기에 나타난 결과로 사료된다. 이는 KBO 리그 내에서 시기에 따른 연봉 결정요인들의 가치가 변화하였음을 시사한다. 또한 타자의 선수 경력은 2012~2016 시즌 KBO 타자모형과 2014~2016 시즌 MLB 타자 모형에서 연봉에 유의한 영향을 미치는 것으로 나타났다. 이러한 결과는 선수의 경력이 타자의 연봉에 통계적으로 유의한 영향을 미친다는 Horowitz 와 Zappe(1998)의 연구와 박승현(2008)의 연구 결과와 일치한다.

또한 2014~2016 시즌 MLB 타자의 연봉 결정요인 분석 결과에서는 다양한 포지션에서 유의성을 나타냈지만, 한국프로야구리그 타자의 연봉 결정요인에서는 소수의 포지션에서만 유의성을 보이고 있다. 이와 같은 결과는 한국프로야구 타자의 연봉을 책정함에 있어 포지션 구분에 따른 선수의 가치 평가는 통계적으로 크지 않았음을 의미한다. 이에 대하여 추후 한국프로야구 현장에서 타자의 가치 평가를 진행한다면 수비 능력, 승리에 대한 수비 관여도 등을 측정 및 평가하여 추가적으로 고려되어야 할 부분으로 사료된다.

## 2. 투수

본 연구에서는 프로야구 투수 연봉 결정요인을 분석하기 위하여 2012~2016시즌 KBO 투수, 2004~2006시즌 KBO 투수, 2014~2016시즌 MLB 투수를 대상으로 세 개의 회귀모형을 설정하여 분석을 진행하였다. 세 모형은 공통적으로 'Allstar'와 'Career'변수에서 연봉과 유의한 선형관계를 보였다. 올스타전 출전횟수는 선수의 인기를 반영할 수 있는 좋은 기준이기예(이군호, 2000) 팬이 없으면 존재 할 수 없는 프로야구의 특성상 각 리그의 구단에서는 이를 반영하여 연봉을 책정한 것으로 보인다.

본 연구의 프로야구 투수 연봉 결정요인 분석에서 가장 주목해야 할 결과는 2012~2016시즌 KBO 투수 모형에서는 선수의 경기력을 나타내는 변수보다 선수의 승리 관련 기록이 연봉에 더 영향을 미치는 것으로 나타났다는 것이다. 2014~2016시즌 MLB 투수 모형과 2004~2006시즌 KBO 투수 모형에서는 투수의 투구 능력에 대한 변수가 연봉에 유의한 영향을 미친다는 결과를 나타냈지만, 2017년 기준 최근 5년의 KBO 투수 연봉 결정요인 분석에서는 이와 반대되는 결과가 도출되었다. 이는 이장영과 강효민(2001)의 연구에서 2000년 KBO리그 투수들은 자책점 보다 승리횟수가 연봉과 더 유의한 관계를 나타내었던 결과와 맥락을 같이한다. 또한 2017년 기준 최근 5년동안의 KBO 리그에서는 투수의 가치를 평가함에 있어 투수 개인 능력보다, 팀의 경기력이 반영될 수 있는 승리

관련 기록이 통계적으로 더 높게 반영되어 평가되고 있음을 시사한다. 후속연구에서는 추가적인 검증을 통해서 위와 같은 가치 평가의 타당성에 대한 검토가 필요할 것으로 사료된다.

### 3. 학문적·실무적 시사점

본 연구는 한국프로야구리그와 메이저리그 선수의 가치 평가 요인에 대한 탐색을 통하여 두 리그의 연봉 결정요인을 비교 분석하였다.

먼저 본 연구를 통한 학문적 시사점은 다음과 같다. 첫째, 한국프로야구 선수의 가치 평가에 대한 연구는 활발히 이루어 졌으나, 다수의 연구에서 경기력의 어떤 요소가 선수의 가치를 높이는지에 대한 연구가 주를 이루었다. 즉 선수의 가치를 평가하는데 있어 경기력과 경기력 이외의 요소를 함께 포함하여 이루어진 연구는 미비했다. 이에 따라서 본 연구에서는 선수의 경기력을 주성분분석을 통하여 변수화한 뒤, 경기력과 경기력 이외의 요소가 연봉에 어떤 영향을 미치는지에 대하여 분석하였다. 이는 경기력만으로 산정하기 어려운 프로야구 선수의 연봉에 대하여, 올스타출전, 올스타팬투표, 선수경력, 좌타우타 구분 등 다양한 요인을 독립변수로 설정하여 검증했다는 점에서 학문적 의미를 찾을 수 있다.

둘째, 단편적인 하나의 세이버메트릭스 지수만으로는 평가하기 어려운 선수의 경기력에 대하여 다양한 세이버메트릭스 지수를 투입하여

주성분분석을 통하여 요인화했다는 점에서 학문적 의미를 갖는다. 마지막으로 특정 시기와 특정 리그가 아닌, 시기에 따른 분석과 리그 구분에 따른 분석을 실시하여 각각의 연봉 결정요인의 차이점을 정리했다는 점에서 학문적 의미를 찾을 수 있다.

다음으로, 본 연구를 통한 실무적 시사점은 다음과 같다. 첫째, 한국프로야구 선수들의 연봉 결정요인 분석을 통하여 매 시즌이 마무리되면 이슈화되는 한국프로야구 연봉 시장에 대하여 분석하고 시기별, 리그별 비교를 통하여 진단하였다는 점에서 실무적 의미를 찾을 수 있다. 둘째, 한국프로야구 연봉 결정요인에 대하여 분석한 결과는 실제 현장에서의 연봉 책정에 있어 기본적인 방향성과 기준점 역할을 할 수 있을 것으로 판단된다. 이에 따라서 선수와 구단의 연봉협상시 타당하고 신뢰로운 연봉 산정을 위한 기초자료로 본 연구의 결과가 반영된다면 실무적으로 의미를 가질 것이다.

## VI. 결론 및 제언

### 1. 결론

본 연구에서는 한국프로야구리그와 메이저리그 선수들의 연봉 결정요인을 탐색하고, 한국프로야구리그의 연봉 시장을 진단하여 한국프로야구리그 선수들의 타당한 연봉 책정을 위한 기초자료를 제시하고자 하였다. 시기에 따른 한국프로야구리그의 연봉 결정요인의 변화와 리그 구분에 따른 연봉 결정요인을 탐색하기 위해 2012~2016시즌 KBO리그, 2004~2006시즌 KBO리그, 2014~2016시즌 MLB리그 선수를 대상으로 다양한 자료를 수집하였다. 수집된 자료는 통계 프로그램 'R Studio 1.0.136 - Mac OS X 10.6+' 을 통하여 주성분분석과 다중회귀분석이 이루어졌다.

타자의 연봉 결정요인 분석 모형에서는 주성분분석을 통하여 산출된 첫 번째 주성분을 경기력 변수로 설정하였고, 투수 모형에서는 첫 번째, 두 번째 주성분을 각각 투수의 투구 능력에 대한 기록, 투수의 승리 관련 기록을 나타내는 독립변수로 설정하였다. 시즌과 리그 구분에 따라 설정한 다중회귀모형은 투수와 타자 각 3개씩 총 여섯 개로 설정하였고, 각각의 모형을 비교 분석하여 도출한 결론은 다음과 같다.

첫째, 한국프로야구리그 타자의 연봉 결정요인은 2012~2016시즌과 2004~2006시즌간 차이를 보였다. 2012~2016시즌 타자의 연봉 결정요인은

선수의 경기력과 FA구분 변수, 올스타출전 횟수/경력 변수에서 다른 독립변수들과 비교하여 상대적으로 큰 t값과 표준화회귀계수를 나타냈다. 반면 2004~2006시즌 KBO리그 타자의 연봉 결정요인에서는 2012~2016 시즌 KBO타자 회귀모형과 비교하여 더 적은 수의 변수에서 연봉과 통계적으로 유의한 관계를 나타냈고, t값과 표준화회귀계수 또한 특정 변수에서만 큰 값을 나타냈다. 특히 2004~2006시즌 KBO리그 타자의 모형에서는 경기력 변수는 유의성을 띠지 않았는데, 이는 본 연구에서 경기력을 세이버메트릭스 지수를 기반으로 산출한 값인 반면 당시에는 선수의 가치를 평가함에 있어 세이버메트릭스 지수의 사용이 미비했기에 나타난 결과로 보인다. 이와 같은 결과는 동일한 리그내에서도 시간의 흐름에 따라 선수의 가치를 평가하는 기준과 각 요인의 중요도가 달라질 수 있음을 시사한다.

또한 2014~2016 시즌 MLB 타자의 연봉 결정요인은 2012~2016 시즌 KBO 타자 모형보다 많은 독립변수에서 유의성을 나타냈고, 특히 선수의 경력 변수에서 큰 t값과 표준화회귀계수를 기록하였다. 이는 한국프로야구 리그와는 달리 세계 최고 수준인 메이저리그에서의 근속년수는 꾸준한 경기력 수준이 유지되어야 가능한 부분이기때 선수의 가치를 평가함에 있어 나타난 결과로 보여진다.

둘째, 한국프로야구리그 투수의 연봉 결정요인은 2012~2016 시즌과 2004~2006시즌 간 차이를 보였다. 2012~2016시즌 KBO 투수의 연봉은 승리 관련 기록이 중요시 되었으나, 2004~2006 시즌 KBO 투수의

연봉에서는 승리에 대한 기록보다 선수의 투구 능력을 나타내는 기록이 중요한 가치 평가 요인으로 나타났다. 하지만 이와 같이 투수의 승리 관련 기록이 중요시 될 경우, 팀의 득점생산력에 따라서 승리에 대한 운이 낮다면 타당한 가치 평가가 이루어지지 않을 수 있기에 프로야구 현장에 검토되어야 할 부분으로 사료된다.

또한 2014~2016시즌 MLB 투수의 회귀모형에서는 투수의 투구 능력에 대한 기록, 선수경력, 올스타 선정 횟수, FA유무, 좌투우투 구분이 연봉에 영향을 미치는 요인으로 나타났다. 특히 한국프로야구리그에서는 나타나지 않았던 좌완투수에 대한 긍정적인 가치 평가가 메이저리그에서는 나타나고 있었다. 이러한 결과는 같은 야구 종목이라도 리그 특성에 따라서 프로선수의 가치 평가가 다르게 이루어질 수 있음을 시사하는 결과이다.

## 2. 제언

본 연구에서는 프로야구 선수의 가치 평가 요인을 탐색하기 위하여 시기에 따른 한국프로야구리그 선수의 연봉 결정요인을 분석하고 추가적으로 2014~2016시즌 MLB 선수의 연봉 결정요인에 대한 분석을 실시하여 한국프로야구와 메이저리그의 연봉 결정요인을 비교 분석하였다. 이에 따른 연구 결과는 한국프로야구리그의 연봉 결정요인에 대하여 통계적으로 진단하였다는 점에서 학문적 의의를 갖는다. 또한 각



회귀모형에서 산출된 유의한 독립변수들의 회귀계수가 한국프로야구 선수들의 타당한 연봉 책정을 위한 기초자료가 된다면 실무적 의의 또한 클 것으로 기대된다.

이와 같은 학문적, 실무적 의의에도 불구하고, 본 연구에서는 자료수집의 한계로 수비능력에 대해서는 변수화하지 못하였고, 선수의 인기를 올스타 팬 투표만으로 한정하였다는 한계점을 갖는다. 또한 각 시즌별 규정타석 및 규정이닝을 충족한 선수만을 분석 대상으로 설정하였기에 리그 전체 선수에게 일반화하기에는 한계가 있을 것으로 사료된다. 이에 따라서 후속연구에서는 선수의 수비능력, 인기도 등 추가적인 다양한 변수의 속성을 잘 드러낼 수 있는 타당한 자료를 수집하여 독립변수로 설정해야 할 것이다. 덧붙여, 분석 대상 또한 규정이닝, 규정타석을 충족시킨 선수에서 더 나아가 대상을 확장하여 연봉 결정요인을 탐색하는 거시적인 연구를 수행해야 할 것이다.

## 참 고 문 헌

- 김기영(1989). *SAS 주성분 분석*. 서울: 自由아카데미.
- 김재희(2011). *R 다변량 통계분석*. 서울: 교우사.
- 김정우, 김기민(2011). 팀내 임금격차가 심화되면 팀 성적이 향상되는가?: 한국프로야구 데이터 분석을 중심으로. *노동정책연구*, 11(1), 1-23.
- 김종환, 고영교, 정희운(2013). 한국프로축구 선수 연봉산정시스템에 관한 연구. *한국체육학회지*, 52(5), 517-526.
- 명왕성, 원영신, 이민규(2016). 의사결정나무 분석을 이용한 국내 프로야구선수들의 연봉 결정요인 분석. *한국사회체육학회지*, 65, 63-77.
- 박성배, 이완영, 전홍권(2016). 한국프로야구 타자 연봉 평가 척도에 영향을 미치는 경기력 변수 분석. *한국사회체육학회지*, 66(-), 55-65.
- 박성희, 강준호(2005). 프로농구선수의 연봉 보완도구로서 상금보상보험에 관한 연구. *체육과학연구*, 16(2), 89-105.
- 박승현(2008). 한국프로야구 타자의 고액연봉에 영향을 미치는 경기력 요인. *한국체육과학회지*, 17(2), 485-494.
- 배재영, 이진목, 이제영(2012). 주성분회귀분석을 이용한 한국프로야구 순위 CSAM(*Communications for Statistical Applications and Methods*), 19(3), 367.
- 배정섭, 박주호, 원도연(2015). DEA 모형을 통한 프로야구 FA선수들의 상대적 효율성 분석. *체육과학연구*, 26(3), 582.
- 승희배, 강기훈(2012). 한국 프로야구 선수들의 경기력과 연봉의 관계 분석. *한국데이터정보과학회지*, 23(2), 285-298.
- 신문선(2002). 한국 프로축구와 프로야구의 연봉제도에 관한 비교 연구. *한국스포츠산업경영학회지*, 7(2), 141.
- 신문선(2004). 한국 프로축구 선수의 연봉산정 모델 개발. *한국스포츠산업경영학회지*, 9(2), 113.
- 양도업, 조은형, 배상우, 정상원(2015). 한국 프로야구 타자의 경기력요인 분석. *한국사회체육학회지*, 60, 305-313.
- 오광모, 이장택(2003). 데이터마이닝을 이용한 한국프로야구 선수들이 연봉에 관한 모형연구. *한국스포츠사회학회지*, 16(2), 295.
- 이군호(2000). 한국 프로야구 선수의 연봉결정에 관한 분석. *미간행 석사학위논문*, 단국대학교 대학원.
- 이만규(2006). 세이버메트릭스를 적용한 프로야구 타자의 경기력과 연봉과의 관계. *석사학위논문*, 국민대 스포츠산업대학원, 서울.
- 이장영(2009). 사회와 문화: 프로 야구선수의 연봉과 소득 불평등의 변화. *社會科學*

- 研究**, 21(-), 87-103.
- 이장영, 강효민(2001). 한국 프로야구 투수의 경기수행과 연봉책정의 관계. **한국스포츠사회학회지**, 14(1), 115.
- 이장영, 최병목(2002). 한국 프로농구선수의 경기능력과 연봉. **한국사회체육학회지**, 17, 879.
- 이장택(2014). 한국프로야구에서 타자능력의 측정. **한국데이터정보과학회지**, 25(2), 349-356.
- 이장택(2015). 한국프로야구 기록들의 장기추세. **한국데이터정보과학회지**, 26(1), 1.
- 이장택(2017). 한국프로야구에서 투수 연봉에 영향을 주는 요인. **한국데이터정보과학회지**, 28(2), 317-326.
- 이제영, 김현규(2016). 한국프로야구에서 타자능력지수 제안-대체선수대비승수(WAR)을 중심으로. **응용통계연구**, 29(7), 1271.
- 최경호(2009). 세이버 메트릭스 소개 및 통계적 측면의 한국프로야구 기록 분석. **사회과학논총**, 25(1), 129.
- 최상진, 박노준(2015). 국내 프로야구선수들의 연봉계약제도에 관한 현안규명연구. **한국웰니스학회지**, 10(1), 31-43.
- 한국야구위원회(2017). **2017 KBO 가이드북 / 한국야구위원회**
- 한필수(2006). 한국 프로 농구선수의 경기실적을 이용한 연봉산정모형 개발. **한국스포츠리서치**, 17(4), 1029.
- 홍종선, 김재영, 신동식(2016). 한국프로야구에서 타자력 지수 제안. **한국데이터정보과학회지**, 27(3), 677-687.
- Baumer, B., & Zimbalist, A.(2013). *The sabermetric revolution: Assessing the growth of analytics in baseball*: University of Pennsylvania Press.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). *Applied multiple regression/correlation analysis for the behavioral sciences*. Routledge.
- Depken, C. A.(2000). Wage disparity and team productivity: evidence from major league baseball. *Economics Letters*, 67(1), 87-92. doi:10.1016/S0165-1765(99)00249-9
- Horowitz, I., & Zappe, C.(1998). Thanks for the memories: baseball veterans' end-of-career salaries. *Managerial and Decision Economics*, 377-382.
- Hotelling, H.(1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6), 417.
- James, B.(1987). *The Bill James Baseball Abstract 1987*: Ballantine Books.
- Jane, W.-J.(2010). Raising salary or redistributing it: A panel analysis of major league baseball. *Economics Letters*, 107(2), 297-299.
- Leeds, M. A., Sakata, S., & Von Allmen, P. (2012). Labor markets and national culture: salary determination in Japanese baseball. *Eastern Economic Journal*, 38(4), 479-494.

- Linoff, G. S., & Berry, M. J. (2011). *Data mining techniques: for marketing, sales, and customer relationship management*: John Wiley & Sons.
- Pearson, K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559–572.
- Richard, B., & David, M. (2003). Performance pay, monitoring environments, and establishment performance. *International Journal of Manpower*, 24(4), 452–471. doi:10.1108/01437720310485933
- Richards, D. G., & Guell, R. C. (1998). Baseball success and the structure of salaries. *Applied Economics Letters*, 5(5), 291–296. doi:10.1080/758524403
- Scully, G. W. (1974). Pay and performance in major league baseball. *The American Economic Review*, 64(6), 915–930.
- Tao, Y.-L., Chuang, H.-L., & Lin, E. S. (2016). Compensation and performance in Major League Baseball: Evidence from salary dispersion and team performance. *International Review of Economics & Finance*, 43, 151–159.
- Todd Jewell, R., & Molina, D. J. (2004). Productive Efficiency and Salary Distribution: The Case of US Major League Baseball. *Scottish Journal of Political Economy*, 51(1), 127–142. doi:10.1111/j.0036-9292.2004.05101008.x

## Abstract

# Exploration of Value Evaluation Factors for Professional Baseball Players

## ∴ Comparison of Salary Determinant Factors in KBO League and MLB

Hwang, Soowoong  
Sport Science of Physical Education  
The Graduate School  
Seoul National University

The purpose of this study is to investigate determinants of the annual salary of baseball players in KBO league and MLB. Based on this investigation, this study diagnoses the current KBO league market and provides a guideline for developing annual salary indicators appropriate for the KBO league.

In order to explore changes of the annual salary determinants in the KBO league, we first compared two groups of seasons: 2012–2016 and 2004–2006. We conducted principal component analysis and multiple regression analysis on records of pitchers and batters among the KBO league players who fulfilled the regulation inning and the regulation batting for each season. The

same analysis was also conducted on MLB players to figure out the difference between the KBO league and MLB. We analyzed records of MLB players in 2014–2016 seasons and compare the annual salary determinants of two baseball leagues. In this process, the principal component analysis is performed to reduce the dimension of the sabermetrics indices, which indicate performance of a player in various aspects. The obtained principal component vectors are used as independent variables of the multiple regression analysis, which provides the determinants of the annual salary.

The results of the study are as follows. First, we have found the annual salary determinants of KBO batters, which show differences between 2012–2016 seasons and 2004–2006 seasons. In 2012–2016 seasons, the important factors for batter's salary are performance, the FA and the number of selections in All-Star Game. On the other hand, in 2004–2006 seasons, the important factors are the FA and the number of selections in All-Star Game while performance of a player doesn't contribute to the annual salary very much. This result shows that even in the same league, the importance of each factor evaluating the value of a player do vary depending on seasons. In addition, for MLB players in 2014–2016 seasons, batter's salary is determined by performance, the career, body weight, the FA and the position, which are almost all variables that we used as input data. In particular, the annual salary of an MLB batter significantly depends on player's career. This shows that a continuous career in MLB is regarded as a reflection of steady performance of a player as MLB is one of the top baseball leagues in the world.

Second, we have found the annual salary determinants of KBO pitchers, which also show difference between 2012–2016 seasons and 2004–2006 seasons. In 2012–2016 seasons, the most important factor for pitcher's salary is the one reflecting winning while, in 2004–2006 seasons, the most important factor is pitching performance. For MLB pitchers in 2014–2016 seasons, the factors contributing the annual salary are pitching performance, the career, the number of selections in All–Star Game, the FA and left/right throws. Especially there is a tendency in MLB that left throw pitchers receive higher salaries, which is not seen in the KBO league. This shows that player's specific feature is evaluated differently depending on leagues.

---

Keywords : KBO, MLB, Salary, Value Evaluation, Principal Component Analysis, Multiple Regression

Student Number : 2016–21636